

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:34:28 ; Search time 103 seconds  
(without alignments)  
5105.057 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVWSPVLEVACKETQLI.....FISSVYAFDTLCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_25Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9514	100.0	1861	3	AAY90350 Drosophil
2	9482	99.7	1954	4	Abb62757 Drosophil
3	541	5.7	934	5	Abp64732 Human pro
4	408.5	4.3	726	4	Aab92933 Human pro
5	408.5	4.3	726	6	Abr48183 Human pro
6	319.5	3.4	526	4	Aab92906 Human pro
7	303.5	3.2	2779	4	Abb62371 Drosophil
8	300.5	3.2	1855	6	Abx47539 Breast ca
9	291	3.1	1742	7	Adb82767 Human pro
10	290.5	3.1	285	4	Aab92905 Human pro
11	284.5	3.0	1805	6	Abu35589 Protein e
12	284.5	3.0	1805	7	Abu36321 Protein e
13	277	2.9	1818	6	Abu36321 Protein e
14	277	2.9	1898	7	Adx48869 Human pro
15	277	2.9	1898	7	Adx48869 Human pro
16	275.5	2.9	2688	4	Aam40883 Human pol
17	271.5	2.9	1975	4	Abb62094 Drosophil
18	270	2.8	2383	5	Abg65631 Human bre
19	269.5	2.8	2663	4	Aam39097 Human pol
20	269	2.8	1525	6	Aab32042 Human ves
21	268	2.8	2482	2	Aar72826 Human mit
22	268	2.8	2482	2	Aaw23996 Human mit
23	265.5	2.8	2633	4	Abg06505 Novel hum
24	264	2.8	2442	3	Aay77575 Human cyt
25	260	2.7	1979	7	Adb75595 Prostate

## ALIGNMENTS

RESULT 1  
AAY90350  
ID AAY90350 standard; protein; 1861 AA.  
XX AC AAY90350;  
XX AC AAY90350;  
DT 04-DEC-2000 (first entry)  
XX DE Drosophila Asp protein sequence.  
XX Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor;  
XX Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor;  
XX Drosophila sp.  
XX WO200052478-A1.  
XX 08-SEP-2000.  
XX 03-MAR-2000; 2000WO-GB0000785.  
XX 04-MAR-1999; 99CB-00005007.  
(UYDU-) UNIV DUNDEE.  
XX Glover DM, Avides MDC;  
XX WPI; 2000-594203/56.  
XX N-PSDB; AAA37761.  
XX Use of Drosophila Asp polypeptide for identifying substances capable of  
disrupting microtubule organizing center integrity and use of the  
identified substances for inhibiting mitosis in tumor cell.  
XX Claim 4; Page 43-44; 51pp; English.  
XX This sequence represents the Drosophila Asp protein. The invention  
relates to the use of Drosophila Asp polypeptide (or its homologue, or  
fragment) capable of stimulating formation and/or maintenance of  
microtubule organising centres (MTOCs), in an assay for identifying a  
substance capable of disrupting MTOCs. Asp polypeptide or its  
homolog is useful for identifying a substance capable of disrupting MTOC  
integrity. Substances identified by the method can be used to inhibit  
mitosis, e.g. in tumour cells

Sequence 1861 AA;

Query Match 100.0%; Score 9514; DB 3; Length 1861;

Best Local Similarity 100.0%; Pred. No. 0;					
Matches 1861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MELVWSPVLEVA	QY	1081	AINIQORWRAK
DB	1	MELVWSPVLEVA	DB	1081	AINIQORWRAK
QY	61	GAGTWMKSVSA	QY	1141	RLRAIMKREORE
DB	61	GAGTWMKSVSA	DB	1141	RLRAIMKREORE
QY	121	POSEPVIISQ	QY	1201	QMRREKRYNHL
DB	121	POSEPVIISQ	DB	1201	QMRREKRYNHL
QY	181	PTLKGNVKSC	QY	1261	QYVHLHREVIT
DB	181	PTLKGNVKSC	DB	1261	QYVHLHREVIT
QY	241	PLMKTITIVH	QY	1321	QYVHLHREVIT
DB	241	PLMKTITIVH	DB	1321	QYVHLHREVIT
QY	301	LSLLSPQTKYA	QY	1381	QYVHLHREVIT
DB	301	LSLLSPQTKYA	DB	1381	QYVHLHREVIT
QY	361	FLFNHSEILA	QY	1441	ATVQARRQRE
DB	361	FLFNHSEILA	DB	1441	ATVQARRQRE
QY	421	SPPKQKQVET	QY	1501	KARQDYQLIO
DB	421	SPPKQKQVET	DB	1501	KARQDYQLIO
QY	481	TSIQPSVKLY	QY	1561	QLLRCSMPGQ
DB	481	TSIQPSVKLY	DB	1561	QLLRCSMPGQ
QY	541	VSPADLDADN	QY	1621	GLLDIRKRIQ
DB	541	VSPADLDADN	DB	1621	GLLDIRKRIQ
QY	601	SEQMRLPCSK	QY	1681	MWCSEFMSTC
DB	601	SEQMRLPCSK	DB	1681	MWCSEFMSTC
QY	661	KIQMSNRD	QY	1741	LLRWCDKDSI
DB	661	KIQMSNRD	DB	1741	LLRWCDKDSI
QY	721	DOAKQKIVKH	QY	1801	NARKPPPMST
DB	721	DOAKQKIVKH	DB	1801	NARKPPPMST
QY	781	LDREDFAFNN	QY	1861	F 1861
DB	781	LDREDFAFNN	DB	1861	F 1861
QY	841	EANPOLGGDI	QY	1921	QKQAAAAASY
DB	841	EANPOLGGDI	DB	1921	QKQAAAAASY
QY	901	RRIRHKLME	QY	1981	QKQAAAAASY
DB	901	RRIRHKLME	DB	1981	QKQAAAAASY
QY	961	SYHSIIITQR	QY	2041	QKQAAAAASY
DB	961	SYHSIIITQR	DB	2041	QKQAAAAASY
QY	1021	QKQAAAAASY	QY	2101	QKQAAAAASY
DB	1021	QKQAAAAASY	DB	2101	QKQAAAAASY

Db	1021	QKQAAAAASYIQMOWRTYQLGRIQRHEFLQRDLIMFVQRRMRKSWMLERQKEFQQLKRA	1081
QY	1081	AINIQORWRAKSMRKCNADYLALRSSVLKVQAVRKATIQMRIDRNHYYSIRKXNVTICLOQ	1140
Db	1081	AINIQORWRAKSMRKCNADYLALRSSVLKVQAVRKATIQMRIDRNHYYSIRKXNVTICLOQ	1140
QY	1141	RLRAIMKREORENYLRLENASILVQKRYMRQOMIQDRNAYLRTKCIINVOQRWRATL	1200
Db	1141	RLRAIMKREORENYLRLENASILVQKRYMRQOMIQDRNAYLRTKCIINVOQRWRATL	1200
QY	1201	QMRREKRYNHLQTTTKRIQIKFRKREMKQRAEFLQLKKVTLVQKRRALLQMPKER	1260
Db	1201	QMRREKRYNHLQTTTKRIQIKFRKREMKQRAEFLQLKKVTLVQKRRALLQMPKER	1260
QY	1261	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1320
Db	1261	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1320
QY	1321	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1320
Db	1321	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1320
QY	1381	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1380
Db	1381	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1380
QY	1381	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1380
Db	1381	QYVHLHREVITKLQRRFHAQKSMFRAKRYGTQAAVSCILQMHWNHLLRKRENSFLQ	1380
QY	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQORREFEMKRRAAVVIQRRFRARCAML	1500
Db	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQORREFEMKRRAAVVIQRRFRARCAML	1500
QY	1501	KARQDYQLIOSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMTEQRCNF	1560
Db	1501	KARQDYQLIOSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMTEQRCNF	1560
QY	1561	QLLRCSMPGQARAGFWARKFQALMTPEMMDLIQKRAKVQRYMRGVLIRRRQKHQ	1620
Db	1561	QLLRCSMPGQARAGFWARKFQALMTPEMMDLIQKRAKVQRYMRGVLIRRRQKHQ	1620
QY	1621	GLLDIRKRIQALQROEAKAVNSVRCKQVAVRFLRGRFTASDALAVLSQLDRLSRTVPHLL	1680
Db	1621	GLLDIRKRIQALQROEAKAVNSVRCKQVAVRFLRGRFTASDALAVLSQLDRLSRTVPHLL	1680
QY	1681	MWCSEFMSTFCYGINAQAIIRSEVDKQLIERSRIILNARYNSTVNTFQEGGLVTTAQM	1740
Db	1681	MWCSEFMSTFCYGINAQAIIRSEVDKQLIERSRIILNARYNSTVNTFQEGGLVTTAQM	1740
QY	1741	LLRWCDKDSIEINTLCTLIWVFAHCPKRRKLIIDHYMTNPEALYVRETCKLVARKEKWKQ	1800
Db	1741	LLRWCDKDSIEINTLCTLIWVFAHCPKRRKLIIDHYMTNPEALYVRETCKLVARKEKWKQ	1800
QY	1801	NARKPPPMSTGRYKSKINFPTCSLPSLEDPFGIIRSPYPTFISSVYAFDTILCKLOIDM	1860
Db	1801	NARKPPPMSTGRYKSKINFPTCSLPSLEDPFGIIRSPYPTFISSVYAFDTILCKLOIDM	1860
QY	1861	F 1861	
Db	1861	F 1861	
RESULT 2			
ABB62757			
ID	ABB62757	standard; protein; 1954 AA.	
XX	ABB62757;		
AC	ABB62757;		
XX	xx		
DT	26-MAR-2002 (first entry)		
XX	xx		
DE	Drosophila melanogaster polypeptide SEQ ID NO 15063.		
XX	xx		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
XX	xx		
OS	Drosophila melanogaster.		

RESULT 2

ABB62757

ID ABB62757 standard; protein; 1954 AA.

XX ABB62757;

XX AC

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 15063.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

OS



XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 XX 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX N-PSDB; ABL06860.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signaling and cell-cell  
 XX interactions.  
 XX Disclosure; SEQ ID NO 15063; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS5773-  
 XX ABB72072). The sequence data for this patent did not form part of the  
 XX printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 1954 AA;  
 Query Match 99.7%; Score 9482; DB 4; Length 1954;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1855; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MELVWSPVLEVACKETLQIDNDRNFRKEVMIILKSKSNQPVKNPKPFPVVGKTLQKSGPT 60  
 DB 94 MELVWSPVLEVACKETLQIDNDRNFRKEVMIILKSKSNQPVKNPKPFPVVGKTLQKSGPT 153  
 QY 61 GAGTKMSVSAVQCKRMSAAAAPPKQWTVTAPSPPAWAHPPPOAPLVEKNVYKT 120  
 DB 154 GAGTKMSVSAVQCKRMSAAAAPPKQWTVTAPSPPAWAHPPPOAPLVEKNVYKT 213  
 QY 121 PQEEPVIYISQPSRLKENLSMPTPGNLLDVINDLRFPTLTETRGKQATIFPDNLAAMP 180  
 DB 214 PQEEPVIYISQPSRLKENLSMPTPGNLLDVINDLRFPTLTETRGKQATIFPDNLAAMP 273  
 QY 181 PTLKGNVKSANDMPRITPDLEDQPATNKTFDVKGSETINISLDTLDCSRIDQPH 240  
 DB 274 PTLKGNVKSANDMPRITPDLEDQPATNKTFDVKGSETINISLDTLDCSRIDQPH 333  
 QY 241 PLNKTTTIVHATHFALACIHEEGSPDPRTKSAIHDLDKDIKLVGSPLRKYSKMD 300  
 DB 334 PLNKTTTIVHATHFALACIHEEGSPDPRTKSAIHDLDKDIKLVGSPLRKYSKMD 393  
 QY 301 LSLSPQTKYATQSGMPNLNEMKIRSIEQNYRYQEQIQIKAKDLNSSSSSEASLAGQOE 360  
 DB 394 LSLSPQTKYATQSGMPNLNEMKIRSIEQNYRYQEQIQIKAKDLNSSSSSEASLAGQOE 453  
 QY 361 FLFNHSEILAQSSRNHLHVGVKSGVPKPNPKRRSHLSFSDAPSNESLYRNETA 420  
 DB 454 FLFNHSEILAQSSRNHLHVGVKSGVPKPNPKRRSHLSFSDAPSNESLYRNETA 513  
 QY 421 SPKQKQVEDTTLPRSAAPANASARSSSAHAPHAQSKFKLAOTMSLMKKATPRKVRD 480  
 DB 514 SPKQKQVEDTTLPRSAAPANASARSSSAHAPHAQSKFKLAOTMSLMKKATPRKVRD 573  
 QY 481 TSIQPSVKLYDSELYMQTCINPDPAATTIDPFLASTWYLDQAVDRHQADFKKWLNAL 540

Db 574 TSIQPSVKLYDSELYMQTCINPDPAATTIDPFLASTWYLDQAVDRHQADFKKWLNAL 633  
 QY 541 VSI PADLADANNKIDVGKLFNEVRNKELVAVPTKEBQSMNLYTKYRLETTLKAAVELFF 600  
 Db 634 VSI PADLADANNKIDVGKLFNEVRNKELVAVPTKEBQSMNLYTKYRLETTLKAAVELFF 693  
 QY 601 SEOMELPCSKVAVYVVKOALIRSRDNLHLDVVMORTILELLCENPLWRLGLEWVFE 660  
 Db 694 SEOMELPCSKVAVYVVKOALIRSRDNLHLDVVMORTILELLCENPLWRLGLEWVFE 753  
 QY 661 KIOMQSNREDIVGLSTFIILNRLFRNKCEBQYSKAYTLTEEYAEITIKKHSLOKILFLLPL 720  
 Db 754 KIOMQSNREDIVGLSTFIILNRLFRNKCEBQYSKAYTLTEEYAEITIKKHSLOKILFLLPL 813  
 QY 721 DOAKOKRIVKGNPCLFVKKSPHKETKIDLLRPFSSSELLANIGDIITRELRGLVLOHROTFF 780  
 Db 814 DOAKOKRIVKGNPCLFVKKSPHKETKIDLLRPFSSSELLANIGDIITRELRGLVLOHROTFF 873  
 QY 781 LDEPDYAFNNLAVDLGDGVLTRVVEVILLRDLTRQLRVPAISRLOIRFNVKLGALG 840  
 Db 874 LDEPDYAFNNLAVDLGDGVLTRVVEVILLRDLTRQLRVPAISRLOIRFNVKLGALG 933  
 QY 841 EANFOLGGDIAAODIVDGHREKTLSSLWQLIYKFRSPKFAAATVLOKWRERHFWLVQ 900  
 Db 934 EANFOLGGDIAAODIVDGHREKTLSSLWQLIYKFRSPKFAAATVLOKWRERHFWLVQ 993  
 QY 901 RRIIRKELMRHRAATVIOAVFRGHOMKYVKLFTERTQAAIILQKTRRYLAOKLYQ 960  
 Db 994 RRIIRKELMRHRAATVIOAVFRGHOMKYVKLFTERTQAAIILQKTRRYLAOKLYQ 1053  
 QY 961 SVHSITTIQWRAOOLGRHORFVELREAAIFLQIRWRRELFKAKLLAAATARLQRS 1020  
 Db 1054 SVHSITTIQWRAOOLGRHORFVELREAAIFLQIRWRRELFKAKLLAAATARLQRS 1113  
 QY 1021 QKQAAAASYIQOMWRTYQIGRIORHEFLRQDLIMFVQRMESKSMLEQKFEQOLKRA 1080  
 Db 1114 QKQAAAASYIQOMWRTYQIGRIORHEFLRQDLIMFVQRMESKSMLEQKFEQOLKRA 1173  
 QY 1081 AINTOORWRAKLSMRKCNADYALRASSVLKQVAYKATIQMRIDRNHYSLRKNVICLOQ 1140  
 Db 1174 AINTOORWRAKLSMRKCNADYALRASSVLKQVAYKATIQMRIDRNHYSLRKNVICLOQ 1233  
 QY 1141 RLRAITMKREORENVLRLNASILVQRYRMEQMIQDRNAVLRTRKCIINVOBRWATL 1200  
 Db 1234 RLRAITMKREORENVLRLNASILVQRYRMEQMIQDRNAVLRTRKCIINVOBRWATL 1293  
 QY 1201 QMRERKNYLHLQTTTKRIQIKFRAKREMKQRAEFLOLKKVTLVVKERRALLQMRKE 1260  
 Db 1294 QMRERKNYLHLQTTTKRIQIKFRAKREMKQRAEFLOLKKVTLVVKERRALLQMRKE 1353  
 QY 1261 QBYLHLREVTIKLQRRFHAQKSMRPMRAKYRGTOAAVSCLOMHWNRHLLRKERSFLQ 1320  
 Db 1354 QBYLHLREVTIKLQRRFHAQKSMRPMRAKYRGTOAAVSCLOMHWNRHLLRKERSFLQ 1413  
 QY 1321 RQAATLORRYEARLNMILKOLKSYAQLKQAAITTIOTRYAKKAMOKOVVLYQKOREAIIK 1380  
 Db 1414 RQAATLORRYEARLNMILKOLKSYAQLKQAAITTIOTRYAKKAMOKOVVLYQKOREAIIK 1473  
 QY 1381 VQRRYRGNLEMRKQIEVYQKQVAVIRLQKWRWSIRDMLCKAGYRRIRLSLSIQKWR 1440  
 Db 1474 VQRRYRGNLEMRKQIEVYQKQVAVIRLQKWRWSIRDMLCKAGYRRIRLSLSIQKWR 1533  
 QY 1441 ATVOARRQREIFLSTIRKVLMOAFIRATILMRQORREFEMKRAAVIQRFRACML 1500  
 Db 1534 ATVOARRQREIFLSTIRKVLMOAFIRATILMRQORREFEMKRAAVIQRFRACML 1593  
 QY 1501 KARQDYQLIQSSVILVQKFRANRSMKQARBEFVOLRTIAVHLQKFKRGLMIQRNCF 1560  
 Db 1594 KARQDYQLIQSSVILVQKFRANRSMKQARBEFVOLRTIAVHLQKFKRGLMIQRNCF 1653  
 QY 1561 QLLRCSMPFGQARAGFMARKRFQALMTPEWMDLLRQKRAAKVIOQRYWRGYLIRRRQKHQ 1620

Db 1654 QLLRCSMPGQARARGFMARKRFQALMTPEMDLIRKRAAKVIQRYWRYLIRRRQKHQ 1713  
 QY 1621 GLDIDRKRIQAQAEAKAVNSVRCKVQAEVFLRGRIASDALAVLSQDLRLSRTVPILL 1680  
 Db 1714 GLDIDRKRIQAQAEAKAVNSVRCKVQAEVFLRGRIASDALAVLSQDLRLSRTVPILL 1773  
 QY 1681 MWCSEFMSTCYGIMQAIRSEVDKQLIERCSRIILMLARYNSTTNTFQGGIVTIAQM 1740  
 Db 1774 MWCSEFMSTCYGIMQAIRSEVDKQLIERCSRIILMLARYNSTTNTFQGGIVTIAQM 1833  
 QY 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKPKRIIHDYMTNPAIYVMTRETKLVARKERMKQ 1800  
 Db 1834 LLRWCDKDSIFNTLCTLIWVFAHCPKPKRIIHDYMTNPAIYVMTRETKLVARKERMKQ 1893  
 QY 1801 NARPPPMSTGRYKSQKINFTPCSLPSLEDFGIIRYSPYTFISSVYAFPTILCKLQIDM 1860  
 Db 1894 NARPPPMSTGRYKSQKINFTPCSLPSLEDFGIIRYSPYTFISSVYAFPTILCKLQIDM 1953  
 QY 1861 F 1861  
 Db 1954 F 1954

## RESULT 3

ABP64732

ID ABP64732 standard; protein; 934 AA.

AC ABP64732;

XX 25-FEB-2003 (first entry)

DT Human protein SEQ ID 392.

XX Human; expressed sequence tag; EST; haematopoietic disorder;

XX central nervous system disease; viral infection;

KW peripheral nervous system disease; non-healing wound; infectious disease;

KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;

KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;

KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;

KW cytostatic; haemostatic; virucide; antibacterial; fungicide;

KW immunostimulant; cerebroprotective.

XX Homo sapiens.

OS W0200259260-A2.

XX 01-AUG-2002.

XX 16-NOV-2001; 2001WO-US042950.

XX 17-NOV-2000; 2000US-00714936.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;

PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-590824/63.

XX N-PSDB; ABQ99318.

XX New isolated polynucleotide, useful in research, diagnostic or

PT therapeutic methods, e.g. preventing or treating disorders involving

PT aberrant protein expression or biological activity.

XX Claim 20; SEQ ID NO 392; 394pp; English.

XX The present invention relates to novel human coding sequences (ABQ99268-

CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in

CC therapeutic, diagnostic and research methods. The polynucleotides may be

CC used in the field of molecular biology as hybridisation probes, primers

CC for PCR, for chromosome and gene mapping, for the recombinant production

CC of protein, or in generation of anti-sense DNA or RNA. The

CC polynucleotides are useful in diagnostics as expressed sequence tags

CC (ESTs) for identifying expressed genes or for physical mapping of the  
 CC human genome. The proteins may be used as molecular weight markers, or as  
 CC nutritional sources or supplements. The proteins may be used to maintain  
 CC and expand cell population in a totipotent or pluripotent state  
 CC useful for re-engineering damaged or diseased tissues, transplantation,  
 CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
 CC polynucleotides and proteins are useful for preventing, treating or  
 CC ameliorating disorders involving aberrant protein expression or  
 CC biological activity, e.g. haematopoietic disorders, central/peripheral  
 CC nervous system diseases, mechanical and traumatic disorders, non-healing  
 CC wounds, immune deficiencies and disorders, infectious diseases caused by  
 CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
 CC reactions and conditions, coagulation disorders, or cancer. The  
 CC polynucleotide sequences of the invention were assembled from ESTs  
 CC isolated mainly by sequencing by hybridisation, and in some cases,  
 CC sequences obtained from one or more public databases. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 934 AA;

Query Match 5.7%; Score 541; DB 5; Length 934;  
 Best Local Similarity 22.4%; Pred. No. 9.5e-32;  
 Matches 232; Conservative 188; Mismatches 350; Indels 266; Gaps 38;

QY 982 RORFVELREAAIFLQ---RIWRRLFAKLLAAETARLQSQQAASVYQWWE-TY 1037  
 Db 4 RQLLRKHKASIVIOSTYRMYOYCFYQKL-----QWATKIIEKYRANK 48  
 QY 1038 QLGRIORHEFLRQDLIMFVQREMRSKWSMLEQKEFQQLKRAAINIQWRRAKLSMRKC 1097  
 Db 49 KKQKVFQHNELKKECTCV-----QAGFQDMNKKIQEQHQAIIQKHCKA-FKIRK- 99  
 QY 1098 NADYLALRSSVLKVQ-AYRKATIQMDRIDRNHYSLR-KNVICIQORLAIMKWRQRENY 1155  
 Db 100 --HYLHRAFPVSIQRRYKRLT-----AVRTQAVICISQSYRGG-FKVRKDIQN- 144  
 QY 1156 LRLRNASILVQKRYVRQOQMIQDENVALTRKCIINVQREWRATLQWRERKNVHLQTT 1215  
 Db 145 --MHRATLIQSFYRHRKVD---YETKTAIVVIQNTYRIVRVYRKKKKFLAVQKS 198  
 QY 1216 TKRIQIKFRAKREMKQRAEFQLKKVTLVQVRRRALLQWRKERQBYLHREVTIKLQR 1275  
 Db 199 VRTIQAAFQGMK--VRQKLVNVEEKAAIV---NQSLCCYRSKTKQYEAQSGVMIQE 253  
 QY 1276 RFHAQKSMRFRAKRYGTQAAVSLQMHWRNHLIRKE-----RNSF 1317  
 Db 254 WYKASGLACSQEAEYHSQSRAAVTIQKAFCRMVTRKLETKCAALRTQFFLQWAVYRRF 313  
 QY 1318 LQLRQAAILTQRRYRRLNNIKQLKSAVLQQAATITQTRYAKKAMQKOVLYQKQREA 1377  
 Db 314 VQQRQAAILTQHYFRT-----WQTRKQFLYRKAVALQNHYRFLSAHQGVILQIRSS 369  
 QY 1378 IIVQRRYRGNLEMRKQIEVYQKQRAVIRLQKWRSDMR-LCK-----AGYR-- 1426  
 Db 370 VLIQARSGFIQKRK---FQEIKNSTIKIQAMWRRYRAKKYLCKVKAACKIQAWYRW 425  
 QY 1427 -----RILSSLSIQRKWRATVQARRQREIFLSTI 1456  
 Db 426 RAHKEYIYAVLKVAKIITQGCFTYTKLERTFLNVRASAIITQKRWAILPAKIAHEHFLMIK 485  
 QY 1457 RKVRLMQAFIRATLLMQRRREFEMKRAAVIQRFRARCAMLKARQDYQLIQSSVILV 1516  
 Db 486 R-----HRAACLIQAHYRG-----YKGEQVSLRQKSAALII 516  
 QY 1517 QRKFRANRGMKQARQBFVQLRTIAVHLQKQFRG---KRLMIEQRNCPQLLR----- 1564  
 Db 517 QKVIIRAEAGHERIKYIEFKTSVILQALVRGMLVRKRFLEQRAKIRLLHFTAAAYVHL 576  
 QY 1565 -----CSMFGFOARAGFMARKEF-QALMTPEMMD---- 1593  
 Db 577 NAVRIQRAYKLYLAVNKANKQVNSVICTQRFWRAR-----LOEKRFIQKYSIKKIEHGG 632





RESULT 6  
AAB92906  
ID AAB92906 standard; protein; 526 AA.

AC AAB92906;  
DT 26-JUN-2001 (first entry)  
DE Human protein sequence SEQ ID NO:11533.

Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.  
PN EP1074617-A2.  
PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
PA (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 11533; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention

Sequence 526 AA;

Query Match 3.4%; Score 319.5; DB 4; Length 526;  
Best local similarity 20.5%; Pred. No. 2.9e-15;  
Matches 118; Conservative 92; Mismatches 124; Indels 243; Gaps 16;

805 VELLRLDITLRLVPAISRLQRIENFKLALGALGEANFOL---GGDIAAQDIDVGDHR 860

Db 1 MELLTQNDLSKKLRIPALSRLOQHNVDIVLQVLSKRGIELSDHGNLTLSKDIVDRHR 60  
Qy 861 EKTLSLWQLLYKFRSP-----KPHAAATVQLQKWRR 892  
Db 61 EKTLLRLWKTAFAFOVDISLNLQDKIEAFLEKHTKSIKKTISLLSCHSDLLINKKGR 120  
Qy 893 -----HWLVV----- 898  
Db 121 DSGSFEQYSENIKLMDWVNAVCAYNKVENFTVPSDGRVLCYLIHHYHPCVPLDAI 180  
Qy 899 ----- 898  
Db 181 CQRTTQVTECTQSGVVLNSSSDSLDMSLKAFDHENTSLEYELLENEKNFHLVR 240  
Qy 899 -----IQRRIRHK-----ELMRHRAATVIQAVFRGHQ 926  
Db 241 SAVRDLGGIPAMINHSMSNTIPDEKVITVLSFLCARLLDLRKEIRARLIQTWRKYK 300  
Qy 927 MRKYVKLFKTERTOAAIILQKFTTRYLAOKOLYOSYHSIIITQWRWRAQQLGRQHRQFV 986  
Db 301 LKTDLKRHQ-EREKAAARIQLAVINFLAKQRLRKRVNAALVIQYW----- 345  
Qy 987 ELREAAIFLQIRWRRRLFAKLLAAAFARLORSQKQQAASYYIQWRTYQLGRIOBHE 1045  
Db 346 -----RRVLAQRKLLMLKKEKLEKQNK-AAASLIQGYWRRYS-----TQR 385  
Qy 1047 ELRQDLTMFYQRRMRSKWMLQKKEFQOLKRAAINIQORWRAKLSMRKCNADYALRS 1106  
Db 386 FLKLYYSIIQSRIR-----MLIATVSYKRYLWATVIQHRWRAVLRKQOQRYEMLKS 441  
Qy 1107 SVLKVQA-YRKATIQRIRDRNHYSLRXNVCLOQRLEA-IMKREORENVYLR-----N 1160  
Db 442 STLIISQSMERK-----WKQRK---MQSQVKATVILQRAFWHLRKQAKSEN 485  
Qy 1161 ASILVQKRYRQMQMIQDRNAVLRTRKCIINVRWR 1197  
Db 486 SALLIQSWRMHKL-----RKYYIRSCVVIQKFR 518

RESULT 7

AAB62371  
ID AAB62371 standard; protein; 2779 AA.

AC AAB62371;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 13905.

Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US009231.

23-MAR-2000; 2000US-0191637F.

11-JUL-2000; 2000US-00614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL06474.

New isolated nucleic acid detection reagent for detecting 1000 or more PT genes from Drosophila and for elucidating cell signalling and cell-cell interactions.







Db 1490 NLILELKPRGVAVNL-----IPGLPAILPWCVRHADYLDNDQKVRSLTSTI 1537  
 QY 1592 MDLIROKRAAKVIQR-----YWRGYLIR-----RRQKHQGLL 1623  
 Db 1538 NSI-----KKVLKRGDDFETVSFWLSNTRFLHCLKQYSGEGFMKHNTRQNEHCLT 1591  
 QY 1624 DIKRIAQLOEAKVNSVCKV-OEAVRFLRG---RFIASDAL----- 1663  
 Db 1592 NF-DLAEYRQ---VLSDLAIQIYQQLVRVLENILOPMIVSGMLHEHTIOQVSGVKPTGL 1646  
 QY 1664 --AVLSQDLRLSRVPHLWMCSEFMSTFCVGINMAQAIRSEVDKOLIERCSRIIIN--- 1717  
 Db 1647 RKRTSSIADEGTYTILDSILRLQNSFHSVMCOHGMDEPILIKQVYKQMFYIICATIANLL 1706  
 QY 1718 -----LARYNSTTV-----NTFOEGG-----LVTIAQML--LRWCDDKSEI 1751  
 Db 1707 RKMCSWSKGMQIRYVNSQLEWLRDKNLMNSGAKETLEPLIQAAQLLQVKKTTDDAEA 1766  
 QY 1752 FNTLCTLI 1759  
 Db 1767 ICSMCNAL 1774

RESULT 9  
 ADB82767  
 ID ADB82767 standard; protein; 1742 AA.  
 XX AC ADB82767;  
 XX AC  
 DT 04-DEC-2003 (first entry)  
 XX Human protein sequence useful for the treatment of cancer (SeqID 1548).  
 DE human; prostate; cancer; cytostatic; gene therapy; vaccine;  
 KW immune response.  
 KW Homo sapiens.  
 OS  
 FN WO2003050236-A2.  
 XX 19-JUN-2003.  
 PD  
 XX 04-SEP-2002; 2002WO-US028214.  
 XX  
 PR 07-DEC-2001; 2001US-00012697.  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSEQ ) HYSEQ INC.  
 XX  
 XX Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;  
 PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;  
 PI Garcia V, Jones LW, Stache-Crain B, Scott EM;  
 XX  
 XX WPI; 2003-513972/48.  
 XX  
 XX New polynucleotides derived from human prostate, useful for modulating  
 PT immune response to prevent or treat cancer.  
 XX  
 XX Claim 18; SEQ ID NO 1548; 188pp; English.  
 XX  
 CC This invention relates to novel isolated polynucleotides of human origin,  
 CC particularly isolated from the human prostate. Specifically, it refers to  
 CC the diagnostics and therapeutics comprising these novel human  
 CC polynucleotides, and includes the derived probes, antisense  
 CC oligonucleotides and antibodies thereof. The identification of these  
 CC human prostate genes that can inhibit tumour growth is useful for  
 CC understanding the progression and nature of complex diseases such as  
 CC cancer, and hence they are important in the drug discovery process. The  
 CC present invention describes these polynucleotides and encoded  
 CC polypeptides as exhibiting cytostatic activity, and through gene therapy  
 CC and/or vaccines they can be used to modulate the immune response for the  
 CC prevention or treatment of cancers, particularly of the prostate, but

CC also for breast, lung and colon cancer. This polypeptide sequence is a  
 CC human protein sequence useful for the treatment of cancer, used in an  
 CC exemplification of the invention. NOTE: These sequences are not given in  
 CC the specification but are provided on the WIPO website.  
 XX  
 SQ Sequence 1742 AA;  
 Query Match 3.1%; Score 291; DB 7; Length 1742;  
 Best Local Similarity 17.8%; Pred. No. 2.4e-12;  
 Matches 292; Conservative 275; Mismatches 513; Indels 560; Gaps 77;  
 QY 415 NETVAISPPKKQKQVEDTTLPRSAAPANASARSSAHAWPHAQSKKFKLAQTSMLMKPAT 474  
 Db 382 SETV-----VKPMTRPQAVNARDALAKKIYAHFLDFIVE---RINQALQSGKQHT 429  
 QY 475 PRKVRDTSIQPSVKLYDSLEYMOTCIN----- 501  
 Db 430 FIGVLIDYI---GEFTFDVNSFEQFCINAYANEKLOQQNMHVFKLEQBEYMKEDIPWTLLD 486  
 QY 502 -PDPPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNALVSIIPADLDADLNKIDVGKL 560  
 Db 487 FYDNQPVIDILIEAKMGILELLEDECLPHGTD-ENWLQKL-----YNNFVNRNPL 535  
 QY 561 FNEVR--NKLIV---APTKEEQSMNYLTKVR-----LETLEKAAVEL---FFSEQMR 605  
 Db 536 FEKPRMNTSFVIQHAFADKVEYKCEGFLEKRNDRVTYDMLVEILRASKFHLCANFQENPT 595  
 QY 606 LPCSKVAVVYNKQALRISDRNLHLDVMQRTILELLLCFNPLMLRLGLEVVFGKIQMQ 665  
 Db 596 PP-----SP-----FGSMITVK 607  
 QY 666 SNRDIVGLSTFLNRLFRNKCEQYSKAYTL-----TEEYARTKKHSLQKILFLLP 718  
 Db 608 SAKQVIRPN---SKHFRTTVGSKFRSSLYLLMETLNATTPHYVVRCKIPNDEK-----LP 658  
 QY 719 PLDAQOKRIVKHNPCLFVKKSPHKEYDILLRFSSELLANIGITRELRLGVYLQHRQ 778  
 Db 659 F--EFDKRIIVQQLRACGVLET-----IRISAQSYPS-----RW 690  
 QY 779 TFLDEFDYAFNNLAVDLRDGVRLTR-----VVEVILLRDDLTQLRVPATSRQLR 828  
 Db 691 TYIEFYS-----RYGILMTYKQELSFSDEKKEVCVLRH----- 723  
 QY 829 IFNVKLALGALGEAN-FOLGGDIAAQDIVDGHREKTLISLLMQLIY--KFRSPKFAAATV 885  
 Db 724 -----LIQDSNOYQFG-----KTKIFFRAGQVAYLEKRLDKLRQSCVM 762  
 QY 886 LQKWWRRHLHVVIQRRIPHKELMERHRAATVIOAVFSGHO-MRKYVKLFKTRTQAAII 944  
 Db 763 VQK-HMRGWL-----QRKKFLERRAAALIIOQYFRGQQTVRKAITAVALKAEAAALI 813  
 QY 945 LQKFTRRYLAOKLYQSYH-STIITQWRWA-----QQLGRQH-----R 982  
 Db 814 LQKHCRGYLV-RSLYQLIRMATITMWAYSRGFLARRRYKMLKEHKAIVILQYARAWLAR 872  
 QY 983 QRFVELREAAIFLQRIWRRRLFAKLI-----LAAATFARLQSQKQQA 1026  
 Db 873 RRFQSIIRFVLNIQLTYRVQRLQKLEKQNKENHGLVEKLTSLAALRAGDVEKI QKLEAB 932  
 QY 1027 ASYIQONQRTVQ-----LGRIORH--BFLQRDLIMFVQRRMRSKWMLQOR 1071  
 Db 933 LEKAAATHRNVEEKGRYRDVAVEEKLAKLQHNSELETQKEQI---QLKQKTELKKEK 989  
 QY 1072 KEFQOLKAAANIQQWRRAKLSMRKCNADYALRSSLVKVQAY-----R 1115  
 Db 990 MD-NLTQQLFDDVQKEEQRMLEK-----SKELKTQDYEQIQLSKEEIKALKDE 1039  
 QY 1116 KATIOQRIDRNHYIS--LRKNVICIQORLAIMKWRBORENVRLRNASILIYQK-----R 1168  
 Db 1040 KMQLQHLVEGEHVTSDGLKAEVARLSKVTKTISEFEKEIE---LLQAKIDIVEKHVQSQK 1096  
 QY 1169 YMRQOM-----IQDENVLIRKCIINQV-----RRWRATLQMRERKNYL 1210  
 :|||:|



ABU35589  
ID ABU35589 standard; protein; 1805 AA.  
XX AC ABU35589;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #21116.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Mycoplasma genitalium.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR N-PSDB; ACA39459.  
XX WPI; 2003-029926/02.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 63513; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX CC Sequence 1805 AA;

Query Match 3.0%; Score 284.5; DB 6; Length 1805;  
Best Local Similarity 19.2%; Pred. No. 8e-12;  
Matches 310; Conservative 264; Mismatches 611; Indels 433; Gaps 75;  
Qy 296 ESMKDLSPQTKY-----AIQSMFNLEMKIRSTEONRYIQ-----EQOIQIK 341  
Db 208 ENOKRLISIEYENTYRELVSADNELQNVYENIDQNIQFHQYQTYRDELSQLERKIQLT 267  
Qy 342 AKDINSSSEASLAGQOEFLEHSEILAOSSRFNHEVGKSVKGSFVNPKRRRSHL 401  
Db 268 KQELVKEKSAURVKIDDADFYINAR--LAE-----LDDVAK-----QL 303  
Qy 402 SFSD--APSNESLYRNETVAISPPK----KQYEDTTLPSA-APANASARSSAHAWPH 454  
Db 304 SFQDGIITQNAQHVDEKLVALNKEKDRLNTQKEAFNLRQSALIDINKLOQENELFA-KH 362  
Qy 455 A--QSKFKLAQTWSLMK----KPAIPRKVYDTSIQSVKLYDSELYMOTCINPDPFAAT 508  
Db 363 LEHQONEFEQOKSDSLKLETEYKALQHKINEFKNESATK--SEELNQ-----ERELPEKR 417  
Qy 509 TTIDPEL--ASTWYLDLDE---QAVDRHQADFKKWLNALVSI PADLDADLN-----NKIDV 557  
Db 418 REIDTLLTQASLEHYEHORESSQLLKDKQNEVKOHQFQNLVYAKKELDKERNLLDQKKVDS 477  
Qy 558 GLFNEVRNKELVVAPTPKEQSMNYLTKYRLETLRKAABELFFSEQMR-----LPC 608  
Db 478 EALF---QLKEKVAQERKELEEL-LYLVKKQKQDKENEL-LFFEKQLKQHQADPENELEA 532  
Qy 609 SKVAVYVVKQALAIRS-----DRNLHLDVVMQRTILELLCFNPLWLRGLGVVFGKIQ 663  
Db 533 KQQLFEAKHALE-RSFIKLEDKKDLNTRAQO-----IANEFSQ 571  
Qy 664 MOSNRDVLGLSTFLNLRFRN-KCEQRYSKAVTLTEEYAEITKKHSLQKILFLPFLDQ 722  
Db 572 LKTDKXSADFELMLQNEYNLQEKQKQLQERTYFERNAAVLNRLNQQKRELL-----626  
Qy 723 AKQRIIVKHNPCLVFKKSPHKTKDILLR-FSEELIAN-----IGDITRELRRLGYVL 774  
Db 627 -----QKQETLDQLTKSFEQERLINQREHKELVASVEKQKELKGL 668  
Qy 775 Q-HRQTFLEDFDYAFNNLA---VDLRGVLTRVVEVILARDDLTQLRVPAISRLQRI 830  
Db 669 QDFSQTSLN---ASKNLAEREMAIKFEKEIEATEKQLND-----VN 708  
Qy 831 NVKLALGALCEANPOLGCD-----IAAQDIVDGHREKXTLSLLWOLLYKFRSPKPHAAATV 885  
Db 709 NAEVQADIALQNLQSLNQERSELSQNAKQRIADFENDSLKKL-----749  
Qy 886 LQKWRRRHVLHVVIQRRIRHKELM-----RRHRAATVIQAVFRGHQMKRYKVLFTERTQA 941  
Db 750 -----NEVELSLQKELQELQTLLEANQKQHSYQN--QAYFEG-----ELDKLNEKQ 793  
Qy 942 AI--ILQKFTFRVLAQKOLYQSVHSIITIQRWRAQQLGHRQRQFVELREAAIFLQRIW 999  
Db 794 AFLNLRKKQTMVEDAIKQRLSDKHQALNMQO-----AELDRKTH-----ELNNA--FLNHDA 843  
Qy 1000 RRLRFAKLLAAAEATAR---LQRS---QKQAAASYI---QMWQ-----RTYQLGRI 1042  
Db 844 DQKSLQDLATVTKETQKLDLERSALLEKQREFAENVAGFKRHSNKTSLQKIYELTKK 903  
Qy 1043 QRHEFLQRDLIMFVQORMSKVSMLEQRK--EPQQLKRAAINIQQRWRAKLSVRKCNAD 1100  
Db 904 QESQTKETELKIAFSDQLQDYQVFELOKQDEFRQ-----IEAKQRELDKLA-EKNQV 957  
Qy 1101 YLALRSSVLVQAVRKATIQRID-RNHYSLRKNVICLOQLRAIMKMBEQ-----1151  
Db 958 KLELDNRFQALQKQDTVQAQLELEREQOLNLEQTAFQANESILKQRELTQKIQAF 1017  
Qy 1152 -----RENYLRLENASILVQKRYRMQOMIQDRNAYLTRKCCIINVOQRWRATLQWRE 1205  
Db 1018 HYELKKRNQFLALKGKELFAKEQDQQRKD--QEIN-----WR-----FKQF 1056  
Qy 1206 RKNYLHLQTTKR-----IQIKFAKEMKKQRAEFLQLKKVTLVQKRRAL 1253







CC trichohyalin protein is useful for forming a proteinaceous gel which may  
 XX then be used for healing wounds, or in breast implants  
 SQ Sequence 1898 AA;

Query Match 2.9%; Score 277; DB 2; Length 1898;  
 Best Local Similarity 21.7%; Pred. No. 3.2e-11;  
 Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

QY 891 RRHLVHVVIQRIIPKHLRHRRAATVIOAVFRGHOMRKVKVLFKTRTQAAILQKFT 950  
 DB 648 RRHLLKSEQERHRHQLRQERREQRRLKREBEERLEQRLKREHE-----ERRE 701  
 QY 951 RYLAOKOLYOSYHIT-IQW-WR-----AQQLG----- 978  
 DB 702 QELAESEQERERIKSRIPKQWQLESEADARQSVLLLEAPQAGRAEAPQEQEKRRR 761  
 QY 979 -----RQHRQRFVELREAAIFLQRIW-----RRRLPAKLLAAETARLORS 1020  
 DB 762 SELQWQEEERAHQQOEEQORRDTWQ--WQAEKSEKRGRLSARPLREQRERQRLAE 819  
 QY 1021 QKQQAASYYIQMWRVYQIGRIQRH-----EFLQRDLIMFVQRRMSKWSMLQ--RKE 1073  
 DB 820 ERQREORFLPEEKEKQGRQRREREKELQFLSEEQ---QRERAQQLQEEEDGLQE 876  
 QY 1074 FOQLKRAAINIQWRRAKLSMRKCNADYLALRSSVLKVQYRKATIQMRIDRNHYSLRK 1133  
 DB 877 DQERRQERQRRDQKRWQLEERKRRH-----TLYAKPALQEQ-----LRX 918  
 QY 1134 NVICLOQLRAIMKWRQRENYLRLNASILVKRYRMQMIQDRNAYLTRKCIINVQ 1193  
 DB 919 EQQLLOEEBEL--QREERKRRQEQ-----RQYREEOLOQEEBQLLEER--EKR 968  
 QY 1194 RWRATIQMRERKNYLHQTTRKI-----QIKFRAKREMKQARFLOIKKVLVVOX 1248  
 DB 969 RQREERQYKDKK-----LOKEQQLGEPKRRQERREKRYREELQOEEBQLLEE 1024  
 QY 1249 RRALLQMRERQRYLHLREVTILQRHFAKSMRPMRAKRYGTQAAVSLQWHWNHL 1308  
 DB 1025 RE-----KRRQEW--ERQYRKDELQOEEBQLLEERKRR-----LQERREQY- 1067  
 QY 1309 LRKRNSFLQROAAITLQRYRPARLNMILKLSYALQKQAITI-----QTRYAKKA 1363  
 DB 1068 ---REEELOQEEBQLLEERTRRQERLERQYRKKEELOQEEBQLLEEFKRRQERE 1124  
 QY 1364 MQ-KQVLYQKORAITKQRYVGNLEMRKQI---EVYQKQOAVIRLQ-----KW- 1411  
 DB 1125 RQREEELOQEEBQLLEERKRRQERLEQYREBELQKQYRDEQSDSLKWQ 1184  
 QY 1412 W-----RSIRDMRL-CKA-----GYRRI-----RLSSLSTIQ-----RKW 1439  
 DB 1185 WEPEKENAVRDNKYCKGRENEQFQLEDQVRDQSQDLOHLLGQEQERDRBQRRRW 1244  
 QY 1440 RATVQARR-----OREFLSTIRKVLQMOPFIRATILMQORRFEMKRRAAVQIR 1491  
 DB 1245 Q---QANRHPFEEQLEERQEKAKRDRKSQ---EBQQLLEEREB---KRRQET-DR 1294  
 QY 1492 RFRARCAMLKARQDYQLIQSVILVQKFRANRSM--KOARQEVQVLRITAVHLQOKF-- 1547  
 DB 1295 KFREEEQLLQEREBQPLLRQ---RDKRFREELLHQEQKFLFEEQRLEERERFLK 1351  
 QY 1548 RGKRLMTBQRCFQLRCSMPGFGQARAGFMARKRFOALMTPEMDLIQKRAKVIQRY 1607  
 DB 1352 EQQLRLER---EQLR-----QDRDKF-----REEEQQLSRQERDKRREE 1392  
 QY 1608 WRGVLTERRQKHQGLDIRKRIALQROE 1635  
 DB 1393 QQ---VARQERKFLSEEQ---QLRQE 1414

RESULT 15  
 ADD48869  
 ID ADD48869 standard; protein; 1898 AA.

XX AC ADD48869;  
 XX DT 29-JAN-2004 (first entry)  
 XX DE Human Protein AAA65582, SEQ ID NO 14580.  
 XX KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX OS Homo sapiens.  
 XX PN WO2003016475-A2.  
 XX PD 27-FEB-2003.  
 XX PF 14-AUG-2002; 2002WO-US025765.  
 XX PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX PF; 2003-268312/26.  
 DR GENBANK; AAA65582.  
 XX PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 PS Claim 1; Page; 1017pp; English.  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
 CC injury (CCI) and spared nerve injury (SNI)).  
 CC The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 1898 AA;

Query Match 2.9%; Score 277; DB 7; Length 1898;  
 Best Local Similarity 21.7%; Pred. No. 3.2e-11;  
 Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

QY 891 RRHLVHVVIQRIIPKHLRHRRAATVIOAVFRGHOMRKVKVLFKTRTQAAILQKFT 950  
 DB 648 RRHLLKSEQERHRHQLRQERREQRRLKREBEERLEQRLKREHE-----ERRE 701





**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:37:54 ; Search time 36 Seconds  
(without alignments)  
4972.567 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVWSPVLEAVACKETLQI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9514	100.0	1861	T13845	microtubule-associ
2	361.5	3.8	1186	T19957	hypothetical prote
3	324.5	3.4	1088	T05464	hypothetical prote
4	324	3.4	1407	S28589	trichohyalin - rab
5	323.5	3.4	1549	A40691	trichohyalin - she
6	320	3.4	1830	S19188	myosin-V - chicken
7	314	3.3	1909	A45592	liver stage antigen
8	304	3.2	1828	B59254	myosin heavy chain
9	301	3.2	4574	G02520	myosin heavy chain
10	301	3.2	4587	A38638	plectin - human
11	300.5	3.2	1375	T05200	plectin - rat
12	300.5	3.2	1855	A59254	myosin heavy chain
13	300	3.2	1374	D85390	myosin heavy chain
14	294	3.1	4684	A59404	myosin-like protei
15	284.5	3.0	1805	A64224	plectin [imported]
16	277	2.9	1818	S73852	hypothetical prote
17	277	2.9	1898	A45973	hypothetical prote
18	271.5	2.9	1853	A46761	trichohyalin - hum
19	267.5	2.8	2663	S28261	myosin heavy chain
20	264	2.8	2442	T08621	centromere protein
21	260	2.7	2245	T18278	centrosome associ
22	254.5	2.7	1837	T41023	myosin heavy chain
23	252	2.6	1313	T48467	probable nuclear p
24	251.5	2.6	1792	T13939	myosin heavy chain
25	250	2.6	2101	T43214	myosin V - fruit f
26	246.5	2.6	2022	T43214	nuclear mitotic ap
27	244	2.6	1846	A59289	ovtl protein - nem
28	243.5	2.6	812	A53016	myr 6, unconvectio
29	243.5	2.6	1410	A57013	myosin heavy chain
					early endosome ant

## ALIGNMENTS

### RESULT 1

T13845

microtubule-associated protein - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000

C;Accession: T13845

R;Saunders, R.D.; Avides, M.C.; Howard, T.; Gonzalez, C.; Glover, D.M.

J. Cell Biol. 137, 881-890, 1997

A;Title: The Drosophila gene abnormal spindle encodes a microtubule-associated protein ti

A;Reference number: Z17792; MUID:97296495; PMID:9151690

A;Accession: T13845

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1861 <SAU>

A;Cross-references: EMBL:U95171; NID:g1930121; PID:g1930122; PIDN:AAB51540.1

C;Genetics:

A;Gene: asp

A;Cross-references: FlyBase:Fgn0000140

C;Function:

A;Description: is required for the normal function of the mitotic spindle

Query Match	100.0%;	Score 9514;	DB 2;	Length 1861;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1861;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELVWSPVLEAVACKETLQIDNENFRKEVWILKSKSNQPVKNPKFPVTVGKTQLQKSPT	60	
Db	1	MELVWSPVLEAVACKETLQIDNENFRKEVWILKSKSNQPVKNPKFPVTVGKTQLQKSPT	60	
QY	61	GAGTKMKSIVSAAVQCKRMASAAAAPPKQTVRVTAPSRPAWAHPPQAPLVEKNVYKT	120	
Db	61	GAGTKMKSIVSAAVQCKRMASAAAAPPKQTVRVTAPSRPAWAHPPQAPLVEKNVYKT	120	
QY	121	POEEPVYLSQPSRSLKENLSPMTCNLVDINDLFTPLTETRGKQATIPFDNLAAPT	180	
Db	121	POEEPVYLSQPSRSLKENLSPMTCNLVDINDLFTPLTETRGKQATIPFDNLAAPT	180	
QY	181	PTLKNVKSANDMEPRITPDDLEDQATNKTFDVKHSETINISLDTDCSRIDGQPHIT	240	
Db	181	PTLKNVKSANDMEPRITPDDLEDQATNKTFDVKHSETINISLDTDCSRIDGQPHIT	240	
QY	241	PLNKTITTVHATHRALACIHEERGSPSPRTPTKSAIHDLRDKIKVGLSPLRKYSKMD	300	
Db	241	PLNKTITTVHATHRALACIHEERGSPSPRTPTKSAIHDLRDKIKVGLSPLRKYSKMD	300	
QY	301	LSLLSPQTKYAIQSGMPNLNEMKIRSIQNYQEQOIQIKAKDLNLSSSSSSEASLAGQOE	360	
Db	301	LSLLSPQTKYAIQSGMPNLNEMKIRSIQNYQEQOIQIKAKDLNLSSSSSSEASLAGQOE	360	
QY	361	FLFNHSEILAQSSRFNLHEVGKSVKQSPVKNPKRRSHLSFSDAPSNESIYRNETVAI	420	
Db	361	FLFNHSEILAQSSRFNLHEVGKSVKQSPVKNPKRRSHLSFSDAPSNESIYRNETVAI	420	

QY 421 SPKKQKQVEDTTILPRSAAPANASARSSSAHAWPHAQSKFKPLAQTMSLMKKPATPRKYRD 480  
 DB 421 SPKKQKQVEDTTILPRSAAPANASARSSSAHAWPHAQSKFKPLAQTMSLMKKPATPRKYRD 480  
 QY 481 TSIPQSVKLYDSELYMOTCINPDPPAATTTIDPELASTMYLDEQAVDRHQADFKKWLNAL 540  
 DB 481 TSIPQSVKLYDSELYMOTCINPDPPAATTTIDPELASTMYLDEQAVDRHQADFKKWLNAL 540  
 QY 541 VSPADLDADLNNKIDVGLKFNERNKELVAPTKBEQSMNYLTKYRLTETLKAQAVELFF 600  
 DB 541 VSPADLDADLNNKIDVGLKFNERNKELVAPTKBEQSMNYLTKYRLTETLKAQAVELFF 600  
 QY 601 SQOMRLPCSKVAVYNNKQALRIRSDENHLVDVMORTILELLCNPLWRLGLVYVGE 660  
 DB 601 SQOMRLPCSKVAVYNNKQALRIRSDENHLVDVMORTILELLCNPLWRLGLVYVGE 660  
 QY 661 KIOMQSNRDI VGLSTFILNRLFNKCEBQYSKAVTLTEEYAEITIKKHSLOKILFLLPFL 720  
 DB 661 KIOMQSNRDI VGLSTFILNRLFNKCEBQYSKAVTLTEEYAEITIKKHSLOKILFLLPFL 720  
 QY 721 DOAKOKRIYKHNPCLVFKSPKETHKIDILLRPSSELLANIGDITRELRLGHVLOHROTFF 780  
 DB 721 DOAKOKRIYKHNPCLVFKSPKETHKIDILLRPSSELLANIGDITRELRLGHVLOHROTFF 780  
 QY 781 LDEFYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPAISRLQIRFNVKLALGALG 840  
 DB 781 LDEFYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPAISRLQIRFNVKLALGALG 840  
 QY 841 EANFOLGGDIAAQDIDVGHREKTLISLMOLIYKFRSPKPHAAATVLOKWRHMLHVVIQ 900  
 DB 841 EANFOLGGDIAAQDIDVGHREKTLISLMOLIYKFRSPKPHAAATVLOKWRHMLHVVIQ 900  
 QY 901 RRIHKELMRRRAATVIOAVRFGHOMKRYVKLFTERTQAAIILLOKETRYLQAKOLYO 960  
 DB 901 RRIHKELMRRRAATVIOAVRFGHOMKRYVKLFTERTQAAIILLOKETRYLQAKOLYO 960  
 QY 961 SYHSIITIQWRRAQOQGRHQRQFVELREAAIFLQIRWRRLFAKLLAAETARLQRS 1020  
 DB 961 SYHSIITIQWRRAQOQGRHQRQFVELREAAIFLQIRWRRLFAKLLAAETARLQRS 1020  
 QY 1021 OKQOAAAASYIQOWRTYQIGRQREHEFLRDLIMFVORRRSKWSMLEQKEFKQOLKRA 1080  
 DB 1021 OKQOAAAASYIQOWRTYQIGRQREHEFLRDLIMFVORRRSKWSMLEQKEFKQOLKRA 1080  
 QY 1081 AINIQQWRRAKLMSRKNADYLALRSSVLKQVAKKATIQMRIDRNHYISLRKNVICLOQ 1140  
 DB 1081 AINIQQWRRAKLMSRKNADYLALRSSVLKQVAKKATIQMRIDRNHYISLRKNVICLOQ 1140  
 QY 1141 RLRAIMKQREORENVLRLNNAISILVQKRYMRQMIQDNAYLRTKCIINVORRWATL 1200  
 DB 1141 RLRAIMKQREORENVLRLNNAISILVQKRYMRQMIQDNAYLRTKCIINVORRWATL 1200  
 QY 1201 QMRERKNYLHLOTTTKRIQIFRAKREMKQBAEFLOLKKVTLVQKRRALLQMKRER 1260  
 DB 1201 QMRERKNYLHLOTTTKRIQIFRAKREMKQBAEFLOLKKVTLVQKRRALLQMKRER 1260  
 QY 1261 QBYLHLREVTIKLQRFHQAQSMRPMKAKYRGTOAAVSCLOQWRNHLIRKERNSTFOL 1320  
 DB 1261 QBYLHLREVTIKLQRFHQAQSMRPMKAKYRGTOAAVSCLOQWRNHLIRKERNSTFOL 1320  
 QY 1321 ROAAITLQRYEYARLNNIKOLASVLAQKQALTIQTRVRAKAMQKOVLYOKOREALIK 1380  
 DB 1321 ROAAITLQRYEYARLNNIKOLASVLAQKQALTIQTRVRAKAMQKOVLYOKOREALIK 1380  
 QY 1381 VQRRYRGNLEMRKQIEVYQKQOQAVIRLQKWRRSIRDMRLCAGYRIRLSSLSIORKWR 1440  
 DB 1381 VQRRYRGNLEMRKQIEVYQKQOQAVIRLQKWRRSIRDMRLCAGYRIRLSSLSIORKWR 1440  
 QY 1441 ATVQARROREIFLSTIRKVRMLQATIRATILMROORREFEMKRAAAVIGRRFRACML 1500  
 DB 1441 ATVQARROREIFLSTIRKVRMLQATIRATILMROORREFEMKRAAAVIGRRFRACML 1500

QY 1501 KARQDYQLIOSSVILVORKFRANRSMKQARQBFVOLRTIAVHLQCKFRGKRLMIEQRNCF 1560  
 DB 1501 KARQDYQLIOSSVILVORKFRANRSMKQARQBFVOLRTIAVHLQCKFRGKRLMIEQRNCF 1560  
 QY 1561 QLLRCSMPGFOARARGFMARKRFOALMTPEMDLIRQKRAAKVIQRYWRGYLIRRKQKHQ 1620  
 DB 1561 QLLRCSMPGFOARARGFMARKRFOALMTPEMDLIRQKRAAKVIQRYWRGYLIRRKQKHQ 1620  
 QY 1621 GLLDIRKRIQLRQAEKAVNSVRCKQEAARVFLRGRFIASDALAVLSQDLRLSRTVPHLL 1680  
 DB 1621 GLLDIRKRIQLRQAEKAVNSVRCKQEAARVFLRGRFIASDALAVLSQDLRLSRTVPHLL 1680  
 QY 1681 MWSEFMNSTCYGIMAQAIIESEVDKOLIERCSRIILNLARYNSTVNTFOEGGLVITIAOM 1740  
 DB 1681 MWSEFMNSTCYGIMAQAIIESEVDKOLIERCSRIILNLARYNSTVNTFOEGGLVITIAOM 1740  
 QY 1741 LLRWCDKXDSIFNTLCTLIWVFAHCPKRRKIIHDYMTNPEAIYVWRETCKLVARKEKMQ 1800  
 DB 1741 LLRWCDKXDSIFNTLCTLIWVFAHCPKRRKIIHDYMTNPEAIYVWRETCKLVARKEKMQ 1800  
 QY 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPFGIIRYSPYTFISSVYAFDTILCKLOIDM 1860  
 DB 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPFGIIRYSPYTFISSVYAFDTILCKLOIDM 1860  
 QY 1861 F 1861  
 DB 1861 F 1861

## RESULT 2

T19957

hypothetical protein C45G3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T19957

R:Barlow, K.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19203

A:Accession: T19957

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1186 &lt;N1&gt;

A:Cross-references: EMBL:Z92780; PIDN:CAB07174.1; GSPDB:GN00019; CESP:C45G3.1

A:Experimental source: clone C45G3

C:Genetics:

A:Gene: CESP:C45G3.1

A:Map position: 1

A:Introns: 21/1; 255/2; 363/2; 575/3; 893/3; 1017/2; 1042/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C45G3.1

Query Match 3.8%; Score 361.5; DB 2; Length 1186;

Best Local Similarity 18.6%; Pred. No. 4.1e-12;

Matches 244; Conservative 216; Mismatches 480; Indels 375; Gaps 47;

QY 458 KKPFLAQWLSMKK---PATPKRVRTSIPQSVKLYDSELYMOTCINPDPPAATTTIDPF 514

DB 21 EKLLDQVKSNTKKIDLRATERAFLESS-----PTSMNMTPLNPS-ISSVSQDSFI 71

QY 515 LASTVYLDQAVDRHQADFKKMLNALVSI PADLDADLNNKIDVGLKFNVEV----- 564

DB 72 LS-----YDEKA-NKQIIALATWCNTM-----NELDVSSEMDLGSKACRNIQWMLKK 120

QY 565 RNKELVVAPTKEQSNY---LTKYRLTETIRKAAVELFFSEOMRLPCKSVAVYNNKQALR 621

DB 121 RSTSEVENTQENARRYQRIFEKNDPEVVKCKCKQLDDSGMD---ASIKDLSSKNVYA 177

QY 622 IRSDRNHLVDVMORTILELLCNFNFLWRLGLVYVGEKIQWQSNRDI VGLSTFTELNL 681

DB 178 IRKEHAVYNDIGLQTLILHTFFSFHPANLKTALAEALFNTRIDAQPKHLMKKLSQFFLDLV 237

QY 682 FRN--KCEBQYRSKAY---TLTEEYAEITIKKHSLOKILFLLPDLQAKOKRIYKHNPCFL 736

DB 238 FSNPTLKNKKFAQSGKPIITEAGKALHGHFLSVSMKLMFLIETATHRVIPNLTRIF 297

```

QY 737 VKKSPHKTCDILLRFSSELL-ANIGDITRELRRLGVLLQHQRTFLDEDFYAPNNLAVDL 795
Db 298 TKSSHENCDDVFSSELTKELLATSSATFKAKAFKGVFIPTYSQSFYENDYQAKGFS-DP 356
QY 796 RGVBLTRVVEV--LLRDDLTQLRVPALSRLORFNVKLALGALGEANFOLG---GDI 850
Db 357 SDGLILAKLLETVGMPHQIILLKLRDPAGDRIRKIGNKIVLQEMS-----SLGVPTDNV 412
QY 851 AAQDIVDGHREKTLILLWLIYKFRSPKPHAAATVLOKWRHHLVHVIQRIR-----H 905
Db 413 NAEISVGGKDELSILMAIL-----GVRVAKEQRIKIVTRVSE 450
QY 906 KELMERHRAA-----TVIOAVFEGHQMKVYKLVKFTERTQAAIILQKFTERRVLAQKQ 957
Db 451 ERTPKRRSAVHDDMSSEVLKMKIYGRQWE--IEVMDLDSLDGCLLAKLITFTGTNST 508
QY 958 LYQSVHSIITIORWRAOQLGQHQRFVELREAAFIQOR--IWRRLRFKALLAAETA 1015
Db 509 PIQYDYG---LSLW-----EKVSVABELELCIQGLDQNNALFVKMFL-----E 549
QY 1016 RLORSQKQQAASYIQMOWRTYQLGRIQRHE-----FL-----RQDLIMP--- 1056
Db 550 RLGMIQDLNEKATKIQRWKAY-----VQRKNTPKLYFIVQQLLADSSIPRNRSVSPFSNN 605
QY 1057 ----VQRMRKSWMLEQRKEFQQLKRAAINIQORWRAKLSMRKCNADYIALRSSVLKQV 1112
Db 606 VTFTVPTPRN--NILTERPSLSQIPSS-----RQSDMTFDATFTVGRDSIESNN 655
QY 1113 AYRKA-----TIOMRIDRNHYYSIRKKNVYICLQORLAIMKREORENYLRLENASI 1163
Db 656 KMQKTPLRGTFRKTIAMVIEEDDSNNETVVPSTLKRTTVVRMEHNAEVE----- 707
QY 1164 LVQKYRMRQOMIQDNAYLRTRKCIINVORWRATLQWRREKNYHLQTTTKRIQIKF 1223
Db 708 -----REQDEDDN-----QDKDTVAPSAB-----NLDSPSPDIPLT 740
QY 1224 RAKREMKQRAEFLQKLVLLVQKRRALLQMRKERQBYLHLREVITIKLQRRFHAQSM 1283
Db 741 LSSIPSASQSAIFLODSE-----EASVDGQKIENLETFEIKE 763
QY 1284 RPRKATYGT---QAAVSCLOMHWNNHLLRKRERNFLQLRQAATLQRRYRRLNMIKQ 1340
Db 764 HVPKAEADVGVVLEASDSPVLEGN-----EASVDGQKIENLETFEIKE 808
QY 1341 LKSYAQLKQAATITQTRYAKKAMQKVVLQKQREAIKQRRYRGNLEWRKQLEVQK 1400
Db 809 KTOEDLPSKPMPTQTSQSPLEVEFRMTQEORLEMLFQSLSEQKRFVKTNNLSVSGIE 868
QY 1401 QRCVIRLQKWRSTIRDMRLCKAGYRRIRLSLSIQRKWRATVQARRQREIFLSTIRKVR 1460
Db 869 DDANTPELRRLIRQTRLEK-----RKQGEI-----AR 895
QY 1461 LMQAFIRATLMRQORREFFEMKRAAVVIORRFRACAMLKARQDYQLQSSVILVQRKF 1520
Db 896 KLGNITERNALAVRGGEDSSDSRSDA-----GHDVAIILHGDDSLFENSMLDQK-- 945
QY 1521 RANRSMKQARQEFVLRTIAVHLQKFRGKRLMTEORNCFOLLRCSPMGFOARARGFVAR 1580
Db 946 ----- 945
QY 1581 KRFQALMTPMMDLTROKRAAKVIQRYWRGYLIRROKHQGLDIRKRIAO-----LRQEA 1636
Db 946 ---SOLQNDTEQILENKKAQAAVIOXMTGRGIARRKFOME--ISNIRNMIQVNHILAQED 1001
QY 1637 KAV-----NSVRCKQVQAEVFLGRFIASDALAVL-----SOLDRLSRTVPHLL 1680
Db 1002 EQIGIEEMEDKSVEAKLKCA--LHG--LTNDNLHVHVHAATVDRVTDVPSLL 1052

```

RESULT 3  
T05464

hypothetical protein T805.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jan-2000  
C:Accession: T05464  
R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.P.X.; Schueller  
submitted to the Protein Sequence Database, February 1998  
A:Reference number: Z15417  
A:Accession: T05464  
A:Molecule type: DNA  
A:Residues: 1-1088 <BEV>  
A:Cross-references: EMBL:AL021890  
A:Experimental source: cultivar Columbia; BAC clone T805  
C:Genetics:  
A:Map position: 4  
A:Introns: 236/3; 393/1; 468/3; 490/3; 508/3; 547/3; 580/3; 641/1; 687/1; 752/1; 845/3;  
A:Note: T805.30  
C:Superfamily: Arabidopsis thaliana hypothetical protein T805.30

Query Match 3.4%; Score 324.5; DB 2; Length 1088;  
Best Local Similarity 18.9%; Pred. No. 4.1e-10;  
Matches 219; Conservative 169; Mismatches 387; Indels 383; Gaps 42;

```

QY 388 SPVKPHKRSHLSFSDAPSNEISLYRNETVAISPPKKQKQVEDTTLPSSAAPANASARS 447
Db 12 APPRNPASSLLSDISNFKTPTRTSVV-NSNISKSPYHPFTASKQTPKSS--SSNFRFRS 68
QY 448 SAHANPHAQSS-----KKFKLAQTMSLMKKPPATPRKVRDTSIOPSVKLYDSELYMOTC 499
Db 69 MVHSYASRSKSVSTSSRRLLKAPFLOQSSRSKAEATKEK----- 106
QY 500 INPDFFAATTIDPFLASTWYLDQAVDRHQADFKKWLNALVSIADLDDAD-LNNKIDVG 558
Db 107 -NLRSLAKSLTV-----WLNFLPENCGCDFFENESGVG 141
QY 559 KLFEVNR-----NKLVLVAPT-----KEEQMNVILTKYRLETLRKA 594
Db 142 NLGCKRDSGEALGNSKSGVDTWMSRPKRLNLGCGEKKHSEIDSSITGSKYSTLR-- 199
QY 595 AVELFFSQMRPLPCS-----KVVVYNKQALRIRSDMLH 629
Db 200 -----ESLRDVCSLDDLKQRMOPHLSIGSKCEIFDVMTRVSKNIDEGRIKWKQCPVL 252
QY 630 LDVVMQRTILELLCFNPLWRLRLGLEVVG-----EKIQMSNRDIVGLSTFILNRLR- 683
Db 253 TDFGKKEKAIKALMSYNQVWRLGLYIIFGDSFLSDSEVNSDQEMAFKMWISQFFSH 312
QY 684 -----NKCEBQRYSKAVTLTEEYAEITIKHSLOKILFLPFLDQAKQRIYK- 730
Db 313 DGLARSVYVYNKWEGLYRPG-----YYEALGNVILKRILLVLDVDRAKSQSCLSLXVG 366
QY 731 -----HNPCLFVKKSPHKEKTDILLRFSSELLANIGDITRELRRLGVVLOHQRTFLDE 784
Db 367 IDGIGGSPIMFSEKSSIKSS-----HQLICGSEF 396
QY 785 DYAFNNLAVDLRDGVRLTRVVEVILLRDLRQLRVPALSRLORFNVKLALGALGEANF 844
Db 397 D-----LMRVLIKQKVVVPSDNRKKNLANCRIALQYLKDGAV 433
QY 845 QLGGD-----TAAQDIVDGHREKTLILLWLIYKFRSPKPHAAATVLOKWRHHLVHVIQ 900
Db 434 SLKDEGMMITGEDVADGDELITISLLWNIPVHLQLP-----LILN 474
QY 901 RRIHKELMRHRAATVIOAVFRGHQMKVYKLVKFTERTQAAIIL-----QKFTTRY 952
Db 475 GRLLTEIYIK-----VQGVQNNQI-----TWSTPLEMLLNMTQDPGQGEQPSV 519
QY 953 LAQQLQVSYHSIITIORWRA-----QQLGQHQRFVELREAAFIQORIRWRRLFAKL 1008
Db 520 MSNTDYDADVQNFILSQKLTALLGSPFEIGDLEHNAVVSQSVIILLAFSLSLIVKEN 579
QY 1009 I-----AAAEATARLQRSQK-----QAAASYIQMOWRTYQLGRI-----QRH 1045
Db 580 LLDFTLAGLNTSVFSQELNFHKLKLLCSCCOVQKRYRIRISCSSESAVTIEPDPRENG 639

```



R.Fietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.  
 J. Cell Biol. 121, 855-865, 1993  
 A>Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding  
 A:Reference number: A40691; MUID:93260018; PMID:7684041  
 A:Accession: A40691  
 A:Molecule type: DNA  
 A:Residues: 1-1549 <FT2>  
 A:Cross-references: EMBL:218361; NID:g295940; PIDN:CAA79165.1; PID:g295941  
 A>Note: sequence extracted from NCBI backbone (NCBI:132511)  
 R.Fietz, M.J.; Presland, R.B.; Rogers, G.E.  
 J. Cell Biol. 110, 427-436, 1990  
 A>Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker  
 A:Reference number: A34209; MUID:90130632; PMID:2298812  
 A:Accession: A34209  
 A:Molecule type: mRNA  
 A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <FT2>  
 A:Cross-references: GB:XS1699; NID:g1827; PIDN:CA35992.1; PID:g1828  
 C:Comment: trichohyalin is a protein of the medulla of the hair and of the inner root sheath  
 Covalent modifications to this protein include conversion of arginine to citrulline and  
 C:Genetics:  
 A:Introns: 46/3  
 A>Note: single copy gene  
 C:Superfamily: trichohyalin; calmodulin repeat homology  
 C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat  
 F:49-81/Domain: calmodulin repeat homology <EF2>  
 F:387-851/Region: 28-residue repeats  
 F:886-1519/Region: 23-residue repeats

Query Match 3.4%; Score 323.5; DB 1; Length 1549;  
 Best Local Similarity 19.3%; Pred. No. 7.6e-10;  
 Matches 280; Conservative 249; Mismatches 504; Indels 421; Gaps 60;

QY 282 RDIKLVGSPRLKYSBMSK-DLSLLSPQTKYALQSGMPLNE-----M 322  
 DB 308 RQKYVSRPHRQEQSRQEQELLERQEQQLSEVQSQEQQLKQEQRLKQEQRYQDNRMW 367  
 QY 323 KIRSTEQNRY-----QEQIQIKADLNSSS-----SEASLAQEQBF 361  
 DB 368 QLEESQRRYTYAKPAQEQVREVEQLRLKEKLRQEQERQYREVELQEEER 427  
 QY 362 LFNHSEIIAQSRFNLHVEGRKSVGSPVKNPHKRSHE-----LSFSDAPNESLYRNE 416  
 DB 428 LQREBEQQLQREER-----EKRRQERQYLEKVELMEEBQEQLEEE 468  
 QY 417 TVAISPPK-KQVDEDTLPSAAPANASARSSAHAWHAQSKPKLAQTWSLMKKPATP 475  
 DB 469 REKRQERQYLEKVEL-REBEQQLQREK-----RQREKQYLEKVELQEEBQ 521  
 QY 476 RKVRTSIPQSVKLYDSLYMOTCINPDPAATTTIDPFLASTMYLDBQAVDRHQADPKK 535  
 DB 522 REERKRQERQYLEKVELQ-----BEEQLQREKREKR 557  
 QY 536 WINALVSTPADLDLNNKIDV---GKLFNEVNTVELVAPTKEQSNYLYTKYRL---E 589  
 DB 558 -----QEREKQYLEKVELQEEQQLQREKQERK-----RQREKQYLEKVELQEE 602  
 QY 590 TLRKAVALFTSEQMLPCSVAVVNVKQALRISDR-----NLHLDVVMQRTILEL-- 641  
 DB 603 QLQREKREKQERQYLEKVELQEEQVQREKREKQERQYLEKREKQERQYLEKREKQ 662  
 QY 642 ---LLCFNPLMLRLGLEVVFGEKIQMQQNRDIVGLSTFLNRLFNKCEQRYSKAYTLT 698  
 DB 663 BEQLLREERKRQERQYLEKVELQEEQ-----LQREERKREKQER-REQYLEK 713  
 QY 699 BEYATIKKSLKILFLPLFDQAKQKRIYVKNPCLFVKKSPHETKDI-----LLRFS 753  
 DB 714 BELQK--QERLQREKEQLQREDEKREKQVREK---YLEBELOQEDRLQREKQLLRED 768  
 QY 754 S-----ELLANIGDITRELRGLGYLQHQRTFLDFDYAFNNLAVDLRDGVRLTRV 804  
 DB 769 REKQYLEKVELQEEQQLQREKQERQYREEL-----LREERLHK 815  
 QY 805 VEVI-----LLRDLRLQRLVFAISRLQRIFNVKLAGALGANF-----QLGQDIAAQ 853

DB 816 EQQLQREBEKRRQRELEQLQEEELQRLDR-----KROFRDDQHQNEVRNS 863  
 QY 854 DIVDGHRE-----KTLSSLWQLIYKPRSPKFFHAAATVLQ-----KWMR-----R 892  
 DB 864 RVYSKHRENKESRQLDDSW-----VRESQFQDLRPLQDEQEKREKREKREKRSRQKRD 918  
 QY 893 HMLVVIQRIHKEMLMRHRAATVIQAVFRGHQWK--YV-----KLFKTRT 939  
 DB 919 QFPABQLLEREQOKETERRDRKFRBEEQLLKQREBKIRYLEEDRKFRBEEQQLRLERE 978  
 QY 940 Q-----AAITLQKFTYRYLAQQLQYQSYHSIIITQWWRQAQQLGRHQKQ--FVE 987  
 DB 979 QQLRQERDRKFRBESRQERDRKFRBEEQLQEE-----REQRURRQERDRKFR 1027  
 QY 988 -----LREAAIFLQRIWRRLP-----AKLLAAAFARLQSQKQAAASYIQMW 1034  
 DB 1028 ERQLQREBEQLRQERDRKFRBEEQLQEEBQLRQERDRKFRBEEQLLQREEL--QLRQERD 1087  
 QY 1035 RYVQLGRQIQRHEFLQRDLIMFVQ-----RMRKSKWMLERKEFQOLKRAAIN 1083  
 DB 1088 ROERNRKFREBQLRREBEQLLQEGEPOLRQDRKFRBEEQLLQREEL--QLRQERD 1145  
 QY 1084 IQORWRKLI-----SMRKCND-----YLALRSSVLKQVAYRKATIQMRIDRNHY 1129  
 DB 1146 RKFREBAQILKEBEQLRQERDRKFRBEEQLLQREELR-----QEREPOLRQERDRK 1201  
 QY 1130 SLRKVVICLOQLRAIMKQREORENYLRNNSIIVQKRYMRQOMIQDNAYLTKCI 1189  
 DB 1202 REBEQLQREKLR-----RQREPOLQER-----DRKFHEEQLQREBQLRQEE-- 1249  
 QY 1190 INVQERWATLQMRERKNYLHLQTTXRIQIKFPAKEMKQBAEFLOKKTIVLVQKR 1249  
 DB 1250 --RDKFREBAQLQREBEQLRQERDR-----KFREBEQLQREEL--QLR-----QER 1296  
 QY 1250 RRLAL-----QMRKEROYLHLREVITIKLQRRHAQ-----KSMRFWRKYRGVQAASVCLQ 1301  
 DB 1297 DKFREBEQLQREBEQLRQEE-----RDKFREBEQLLKESEQLRQERDR-----K 1345  
 QY 1302 MHWNRHLKREKERNFLQRLQAAITLQRRYRARLNMKQLKSYAQLKQAAITQIYRAK 1361  
 DB 1346 PHEKHLRLREEQ--QLRQELE-----GVFSQEQELRAEQEERQRRQQRDR 1393  
 QY 1362 KAMQKQVLYQKREAIIVQRRYRGNIEMKQTEVYQKQAVIRLQKWRKSIDRMLC 1421  
 DB 1394 KPLEBEQQLQREBEKREKRVQEDRKFLQEBEQL--HREBEQLRQEE----- 1440  
 QY 1422 KAGYRRIRLSSISIQKWRATVQARRQREIFLSTIRKVLQMAFTRATLLMRQORREPEM 1481  
 DB 1441 -----LDQYFAEQFAE-----EKRRQEQELQREBQRRQERE-- 1476  
 QY 1482 KRAAVIQRRFRACAMLKARQYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAV 1541  
 DB 1477 -----RKFRBEEQLRQEQEE-----EQKRRQERDVQOSRRQ----- 1507  
 QY 1542 HLQKFRGKRLMIE 1555  
 DB 1508 -VWEEDKGRQVLE 1520

## RESULT 6

S19188

myosin-V - chicken

N/Alternate names: calmodulin-binding protein; myosin I heavy chain, 190K

C/Species: Gallus gallus (chicken)

C/Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text\_change 19-Jan-2001

C/Accession: S19188; A44359; B44359; S29249

R/Espresso, E.M.; Cheney, R.E.; Matteoli, M.; Nascimben, A.A.; De-Camilli, P.V.; Larsc

submitted to the EMBL Data Library, February 1992

A/Description: Complete cDNA coding sequence of chicken brain p190, a calmodulin binding

A/Reference number: S19188

A/Accession: S19188

A/Molecule type: mRNA



A;Residues: 1-1830 <ESP>  
A;Cross-references: EMBL:Z11718  
R;Esprefaco, E.M.; Cheney, R.E.; Matteoli, M.; Nascimento, A.A.; De Camilli, P.V.; Lars  
J. Cell Biol. 119, 1541-1557, 1992  
A;Title: Primary structure and cellular localization of chicken brain myosin-V (p190), a  
A;Reference number: A44359; MUID:93107155; PMID:1469047  
A;Accession: A44359  
A;Molecule type: mRNA  
A;Residues: 1-1688, R', 1690-1830 <ES2>  
A;Experimental source: brain  
A;Note: sequence extracted from NCBI backbone (NCBIN:121153, NCBIPI:121154)  
A;Note: the codon CGC for residue 1689 is inconsistent with the codon AGC for residue 16  
A;Accession: B44359  
A;Molecule type: protein  
A;Residues: 155-164 <ES3>  
R;Sanders, G.; Lichte, B.; Meyer, H.E.; Killmann, M.W.  
FEBS Lett. 311, 295-298, 1992  
A;Title: cDNA encoding the chicken ortholog of the mouse dilute gene product. Sequence c  
A;Reference number: S29249; MUID:93012002; PMID:1383040  
A;Accession: S29249  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1142, 1144-1830 <S&N>  
A;Cross-references: EMBL:X67251; NID:963364; PID:CAA47673.1; PID:963365  
C;Comment: The neck domain comprises six approximately 23-residue tandem repeats; this d  
C;Superfamily: myosin MYO2; myosin motor domain homology  
C;Keywords: actin binding; ATP; brain; coiled coil; nucleotide binding; P-loop; phosphop  
F;72-752/Domain: myosin motor domain homology <MOT>  
F;163-170/Region: nucleotide-binding motif A (P-loop)  
F;645-666/Region: actin binding #status predicted  
F;765-909/Domain: neck #status predicted <NBC>  
F;912-1420/Domain: coiled coil #status predicted <COI>  
F;1421-1830/Domain: carboxyl-terminal domain #status predicted <CTD>  
F;169/Binding site: ATP (Lys) #status predicted  
F;1735/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 3.4%; Score 320; DB 1; Length 1830;  
Best Local Similarity 18.2%; Pred. No. 1.5e-09;  
Matches 350; Conservative 275; Mismatches 659; Indels 640; Gaps 82;

QY	241	PLNKTIVTHATRALACIHEBGPDPPTPTKSAHDLKDKIKVSGPLKYSKMD	300
Db	61	PLRNPDIIVGENDLTALSYLHE-----PAVLHNLK--VRFDISKI-IYTYGIV	106
QY	301	LSLSPQTKVAIQG-----SPNLNEM--KIRSIQNRVYQEQQIKAKDLNSSSS	351
Db	107	LVALNPYEQLPIYGEDIIINAVGQNGMDPHIFAVAEAYQWADERNQSIIVSGESG	166
QY	352	EASLAGOQEFNFHSEILAQSSRFNLHEVGRKSVKSPVKNPKRRSHLSFSDAPNES	411
Db	167	AGKTVSAKAMRYFATVSGSASEANVEE---KVLASNP-----MESIGNAK	210
QY	412	LYRNET-----VALSPKKQKV-----EDTTIPRANAPANASARS-----SAHA	451
Db	211	TTRNDNSRFGKYTEIGDFKRYRIGANMRTYLLKSRVWFQABEERNYHIFYQLCASAA	270
QY	452	WPAQSKFKPLAQTMSLMKKPATP-----RKVRDTSIQPSVKLYDSLYMQTCI----	500
Db	271	LPEFKTLRLGNANFHYTKQGSFVIDGIDAKENWVTRQACTLLIGISDSYQMGIFRILA	330
QY	501	-----NPDPPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWL-----	537
Db	331	GILHLGNVEFASRSDSCAIPKHDPL---TIFCDLMGVD--YEEMAHWLCHRKLATATE	385
QY	538	-----NALVSTPADLDLNNKI--DYGKLFNRYNKLVELVAPTKESQSMNYL	583
Db	386	TYIRPIKSLHAINARDALAKHYANLFWNIVDHVNKALH-----STVKQHSFGVL	436
QY	584	TKYRLTLRKAARELVFFSEQMRLPCKSVAVYVKNQALRISDRNLHL-----DVVMQRTI	638
Db	437	DIYGFETPEINSFQF-----CIN---YANE---KLOQFNNWHYFKLQEEYMKEQI	482
QY	639	LELL-----CFNPLMLRLGLEVVFGKEIKMQSNRDIIVGLSTFILNRLFRNKC-----EE	688

Db	483	PWTLIDFYDNPQPCINLFEAKMGVLDLDECKMPKSGDDTWAQKLYNTHL--NKALFEK	540
QY	689	QRYV--KAYTLTE-----EYAEITKKHSLOKILFLFLPFLDQAKQRI	728
Db	541	PRUSNKAFLIKHPADKVEVQCEGLFKNKDVTVEEQIKVLKSSKFKLLPFLQDEKAI	600
QY	729	-----VKHNPCLFVKKSP-----HKETKDILIRFSELLANIGDITRELRL	770
Db	601	SPTSATSPSGRVPLSRTFVPAKARPQOTSKEHKKTGVGHQFNSLHL-----METL	651
QY	771	GYVLQHQRTFLDFEYAFNNLAVLDGQVLTTRVVEVILLRDLDTQLRVPV-----	822
Db	652	NATTPHYVRCIKENDFKP--PFTFDEKRAVQQLRACGVL-----ETIRISAAGFSRWT	703
QY	823	-----ISRLQRIFNVLALG-----ALGRANFOLGGDIAAQDIVDGHREKTL	865
Db	704	YQEFFRYRVLMQKQVLSDRKQTKCNVLEKLIDKDKYQFG-----KUKIFF	751
QY	866	LLWQLIY--KFRSPKFHAAATVLOKWRHHLHVVIQRRIRHKELMRHRAAATVIOAVR	923
Db	752	RAGQVAYLEKIRADKLRAACIRIQK--TIRGWL-----MRKKYMERRAAIIQRYVR	802
QY	924	CHQMKVVKLFKTERTQAAIILOKFTERYLAQKQ-----LYQSY-----	962
Db	803	GHQARCYATFLR--RTAAIIIOKFORMYVVRKYQCMRDATTALQALLRGYLVNRKYQM	860
QY	963	-----HSIIITIQ--RWRAQALGRQHRQREVELREAAIFLQRIWRRRLFAKILLAAETA	1015
Db	861	MLREHKSIIILQKVRGW---LARVYHRTL---KAIVYLOCCYRMMWAKRELKKLIEA	913
QY	1016	RLORSOKQAAASYIQMWETVYQGRIGRHFELR--ORDLIMFVQRRMRKSWMLRQKE	1073
Db	914	R-----SVERYKHLHIGLENKIMQLQKID-----EQNKE	943
QY	1074	FOOLKRAAINIQORWRAKLSMRKCNADYLAALRSVLKQVAYRATIQMRIDRHHVYSIRK	1133
Db	944	YKSLLEOMNNLEITYSTE--TEKLRSVVERLWS---EEBAKNAT-----NKVLSQE	991
QY	1134	NVICLOQLRAIMKRBORENYLRLNASILVQKRYMRQMI---ODRNAYLFRKCI	1190
Db	992	BIAKRLKELHOTQTEKKTIEW-----ADKYHETELQVSELKEQNTLLKTEKEL	1042
QY	1191	N-----VORRWRAFLQMRERRRNYLHLQTTTKRTOIKF-----	1223
Db	1043	NRRTHDQAKETITMEKKLVEETKQLELDLNDERLRYQNLINFEFSRLEERYDLDKEMNL	1102
QY	1224	-----RAKEMKKORAE-----FLOL-K	1240
Db	1103	MVSIPIKPGHKRTDSTHSSNESEYTFSEITEADPLRMSEQEPSEKAPLDMSLFLKLOK	1162
QY	1241	KVTLVQKRRRALIQMEKBERQEVYHLHREVITIKLQRRFHAQKSMFRKAYRGTOAAVSCL	1300
Db	1163	RVTLEQEKOS--LQDELDRKEQALR-----AKAKEERPPRGABELEYESL	1208
QY	1301	QMHWRNHL--LRKPERNSFLQROAALTQOR-----RYRARNLMKQLKSYAQLK	1348
Db	1209	K---RQBLESENKKLKNELNEL--QKALTETRAPEVTAPGAPAYRVLDDQTSVSEELVR	1264
QY	1349	QAATITQTRVRAKAMQKVLYQKQREAI-----KVORRYRG-----N	1388
Db	1265	KBEVLI---LRSQLVSKQKATQPKEDKNTWDTSTILLEDVQMKDKGEIAQAYILGETVN	1321
QY	1389	LEMKQLEVYQKQQAIVRLQKWWRSIRD--MRLCKAGYRRIRLUSLSIQKRWATVOAR	1446
Db	1322	RLLESQLOSKKSHENEL-----ESLRGEIQLKBEENRQQLLAQMLQPPPEARIEAS	1375
QY	1447	QREIFETSTIKVFLMOAFIRATLIMQORREFPEKRAAVVIQRRFRACAMLKARQDY	1506
Db	1376	LOHEITRLTNENLDMEQ-----LEKQDKTVRLKKQLKVFACKIIEGVQGMENISPG	1429
QY	1507	QLIOSSV--ILVQKFRANRSMKQARQEFVL-----RTIAVHL-----	1543

Db 1430 QIDPIRPNVPRKEDFOGMBLYKKEDBQKLVKNLILSLKPRGVAVNLIPLGPAVILF 1489  
 QY 1544 -----QQRFR-----GKRLMIEQR-----NCFOLLRCMSPGFOA 1572  
 Db 1490 MCVRHADYNDQKVRSLTSTINGIKKVLKRGDDPETVSWLNTCRFLHC----- 1542  
 QY 1573 RARGFMARKRQALMTPE-----MMDLIRQ--KRAKVIQRYWGYLI 1513  
 Db 1543 -LKOYSGEGFMKINTPRONEHCLTNFDLAEYRQVLSDLAIOIYQQLVRVLENIQPMIV 1601  
 QY 1614 RRRQKH--OGL-----LDIRKRIAOLRQBAKAVNSVRCKVQBAVRFLRGFTASDALAV 1665  
 Db 1602 SCMEHETIQGVSGVKPTGKRTSSIADEG----- 1632  
 QY 1666 LSQDLRIASRTVPHLLMCSBPMSTFCYGINAQAIRSVVDKOLIERCSRIILN----- 1717  
 Db 1633 -----TYLDSIIRQLNSPHSVWCQGMGPDLIKQVQMFYIIGAVTLNNLLARKDM 1685  
 QY 1718 -----LARVNSTTV-----NTFOEGG-----LVTIAQML--LRWCDKDSIFNTL 1755  
 Db 1686 CSWSKGMQIRVNSQLSEWLDRKMLNSGAKETLEPLIQAAQLLVKKTDEDAEAIKSM 1745  
 QY 1756 CTLI 1759  
 Db 1746 CNAL 1749

RESULT 7  
 A45592  
 A;Species: Plasmodium falciparum  
 C;Date: 22-Nov-1993 #sequence revision 02-Dec-1994 #text change 09-Jun-2000  
 C;Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592  
 R;Zhu, J.; Hollingdale, M.  
 submitted to the EMBL Data Library, November 1990  
 A;Reference number: S24597  
 A;Accession: S24597  
 A;Molecule type: DNA  
 A;Residues: 1-1909 <ZHU>  
 A;Cross-references: EMBL:X56203; NID:g9915; PID:g9916  
 R;Zhu, J.; Hollingdale, M.R.  
 Mol. Biochem. Parasitol. 48, 223-226, 1991  
 A;Title: Structure of Plasmodium falciparum liver stage antigen-1.  
 A;Reference number: A45592; MUID:92107224; PMID:1840628  
 A;Accession: A45592  
 A;Molecule type: DNA  
 A;Residues: 1-1957638-688;1165-1215;1590-1909 <ZHU>  
 A;Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75014, NC  
 R;Guerin-Marchand, C.; Druiilhe, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudoin,  
 Nature 329, 164-167, 1987  
 A;Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene  
 A;Reference number: S29393; MUID:87315391; PMID:3306406  
 A;Accession: S29393  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 323-387 <GUEL>  
 A;Cross-references: EMBL:M28266  
 R;Guerin-Marchand, C.; Druiilhe, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudoin,  
 submitted to the EMBL Data Library, April 1992  
 A;Description: a liver-stage-sepic antigen of plasmodium falciparum characterized by  
 A;Reference number: S34842  
 A;Accession: S34842  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 323-381, 'HKAT', <GUE2>  
 A;Cross-references: EMBL:M28266  
 A;Note: difference at carboxyl end due to frameshift error  
 C;Comment: This protein is found as flocculent material in the parasitophorous vacuole.  
 C;Superfamily: trichohyalin; calmodulin repeat homology  
 C;Keywords: EF hand  
 F;154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-S-D-L-E-Q-E-R-R)  
 Query Match 3.3%; Score 314; DB 2; Length 1909;

Best Local Similarity 18.3%; Pred. No. 3.4e-09;  
 Matches 302; Conservative 334; Mismatches 733; Indels 278; Gaps 50;

QY 109 QAPLVEKNVYKTPQEPVYISPPQPSLKXENISPMTPGNLLDVINDLRFPLTETRGKQA 168  
 Db 162 QSDLEQERLAKELQEQSDSSEQERLAKELQ-----EQQS 197  
 QY 169 TIFPNLAAMPPTTKGNVKSANDMRPRRTTPDDLEDOPATNKTFFVKHSETNISLDT 228  
 Db 198 DLEQERLA-----KEKLEQEQSDLEQERLAKELQEQSDLEQ-ERRAKEKLEQEQSD 249  
 QY 229 LDCSRIDGQPHPLNKTITVHATHRALACIHESGSPPTPTKSAIHLKDKDKILVG 288  
 Db 250 LEQER-----RAKEKLEQEQSDLEQERLAKELQEQSDLEQER 288  
 QY 289 SPILRYSHMKDLSLLSPQTKYAIQGSMPNINEMKI-----RSIEONRYVQEQOIQ 339  
 Db 289 LAKEKLEQEQSDLE-QERRAKEKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQ 346  
 QY 340 IYAKDINSSSEASLAGOBFNFHSEILA-----QSSRFNLHEVGKSVKSPVKNP 393  
 Db 347 EQSDLEQERLAKELQEQSDLE-QERLAKELQEQSDLEQERLAKELQEQSDLE 404  
 QY 394 HKRRSHELFSFSDAPNESLYRNETVAISPPKKQR--VEDTTIPRSAAAPANASRSSAHA 451  
 Db 405 QERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQER 460  
 QY 452 WPHAQSKAFKLAQTMWSLMKKPATPRKVRDTSIQPSVKLYDSELYMQTCINPDPAATYTI 511  
 Db 461 KEKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 505  
 QY 512 DPFPLASTMYLDQAVDRHQADFKNLNAIVSPADLDADLNKIDVGLFNEVRNKLIV 571  
 Db 506 EGERLAKELQEQSDSEGERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 563  
 QY 572 APTKEEQSMNYLTKYRLETLRKAAYELFFSEQMRLPCKSAVYVYNNKQALRIRSDRNHL 631  
 Db 564 EKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 616  
 QY 632 VVMQRTILELLCFNPLWRLGLGVVFGKIQMSNRDIVGLSTFILNRLFNKCEQR- 690  
 Db 617 LQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 656  
 QY 691 -----YSKAYTLTEBYA-----ETIKKHSIQKILFLLPFLDQAKOKRIVKHNPCLPVKKSP 741  
 Db 657 DLERTKASKETLQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 708  
 QY 742 HKETKDILLRFSSELLANIGDITREIRLGVYVLOHRTFLDEFDYAFNNLAVDLADGVRL 801  
 Db 709 LEQER-----RAKEKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQER 762  
 QY 802 TRVVEVIL--LRDDLTROLRVPAISRLQR-----IFNVKIALGALGEANFOLGGDIAAODI 855  
 Db 763 DRLAKEKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 820  
 QY 856 VDGHREKTLISLWLIYKFRSPKFAAATVLOKWRHRLHVLHYIQRIRHK----- 906  
 Db 821 L-----QEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLE 863  
 QY 907 -----ELMRHRAATVIOAVFRGHQKRVKVLKFTERTQAAIILQKFTRY 952  
 Db 864 QDRLAKEKLEQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 922  
 QY 953 LAOKQLYQSYHSIITIQRWRAQQLGRHQRFRVSLREAAIFLQR-----IWRRLFAKKL 1008  
 Db 923 LQEQER-----DLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLE 974  
 QY 1009 LAAETVARLQRSKQAAASYIQMWRTYQLGRIQRHBEFL-RQDLIM--FVQRNRSKW 1065  
 Db 975 QEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 1029  
 QY 1066 SMLQ-RKEFOOLKRAALNIQQRWAKLSMRKCNADYLALRSSVLKVQAYRKATIQMID 1124



```

1178 DRNAVLRTRKCIINQRRWRATLOMERERKNVYLHLQVTTKRIQIKFR-----1224
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1048 QAKMETETMEKGL-VBETQLELDLNDERLRYQNLNERSLEERVDLLKEEMTLMVHPV 1106
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1225 -----AKRMKKQQAELFLQKKVT-----LVQKRRRALL 1254
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1107 KPGKRTDSTHSNESEYIFSSIEAEMDIPSTPEPSKKVPLDMSLFKLQKRVTELE 1166
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1255 QMKERQEVYHLREVTIKLQRRFHAQKSMRFWAKYRGTCQAASVCLQMWHRNHL--LRKR 1312
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1167 QEKQVQWQBLDRKEQVL-----RSKAKEERPOIRGAELYESLK--ROELESSEKK 1217
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1313 ERNSFLQRLQAAITLQRR-----YRARLNMKQKLSYAOLKQAAITIQRYRA 1360
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1218 LKNELNELRKA--LSEKSAPEVAPGAPAYRVLMELQTSVSEELDVKEEVL----LRS 1271
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1361 KMAQKQVVLQKQREAIL-----KVORRYGNLEMRKQIEVYQKQROA 1404
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1272 QLVQKQKAIQPKDDKNTMTDSTILLEDVQMKDKGEIAQAYIGLKETNRLLE-----1323
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1405 VIRLQKWRISIRD-----MELCKAGYRIRLSSLSIQKWRATVQARRQRIFLST 1455
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1324 -SOLQSRSHENEBALRGETQSLKEENRQOQLLAQNLQPPPEARIEASLQHEITRLT 1382
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1456 TRKVELMQAIFRATILMRQORREFEMKRAAVVIORRFRARCAMLKARQDYQLIQSSV--1513
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1383 NENLDLMEQ-----LEKQDKTVRLKKQKVPFAKKIGELVQOMENISPGQIIDEPIRP 1436
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1514 ILVQKFRANRSMKQARQBFVQLRTTAVHLOKFRGKRLIMIQORNCQFLLRCSMFGQAR 1573
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1437 VNIPRKEKDFQGMLEYKEDEQKLVKNLILELKPGRGAVNL-----IFGLPAY 1484
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1574 ARGEMAR-----KRFQALMTPEMDLIRQKRAKVIR-----YWRGVILR--1614
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1495 ILFMCVRHADYNDQKVRSLTSTINSI-----KKVLKRGDDPFTVSWFSLNTRCPL 1538
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1615 -----RRQKHQGLDTRKRIQALRQBAKAVNSVRKV-QBQAVRFLRG-1655
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1539 HCLQKYSGEHGMKHTSRQNEHCITNF--DLAEVQ--VLSDAIQIYQQLVRVLENI 1593
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1656 --RFTASDAL-----AVISQDLRLSRTVPHLLMWCSEFMSTFCYIM 1695
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1594 LQPMIVSGMLEHETIQGVSGVKPTGLRKRTSIADEGTWTLDSILRLNSFHSVMQCHGM 1653
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1696 AQATRSEVDKOLIERCSRIILN-----LARYNSTIV-----NTFFQGG 1733
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1654 DPelikQVVKQFYIIGAITLNNLLLRKDMCSWKGKQIRVNSQLEWLRDKNLMNSGA 1713
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1734 -----LVITIAQML--LRWCDKDSIFNTLCTLI 1759
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1714 KETLEPLIQAAQLQVKKKTDADAICSMCNAL 1747
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
G02520
protein - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C;Accession: G02520
R;McLellan, W.H.I.; Smith, F.J.D.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01385
A;Accession: G02520
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-4574 <MCL>
A;Cross-references: EMBL:U53204; NID:g1477645; PIDN:AAB05427.1; PID:g1477646
C;Genetics:
A;Gene: pLEC1
C;Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
P;68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

```

```

Query Match      3.2%; Score 301; DB 2; Length 4574;
Best Local Similarity 19.6%; Pred. No. 6.1e-08;
Matches 334; Conservative 276; Mismatches 652; Indels 446; Gaps 66;

QY      161 ETRGKQATIF--PDNLAAMP-----TPTLKGNVKSCANDMRPRRTTPDDLEQDPATNKT 213
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1061 EGLGKGVARSAAEAKVLAPEPSPAAPTILRSELE-----LTLGKLEQVRSLSAI 1110
QY      214 FDKHSETTINISLDTDCSRIDGQPHTPLNKTTIVHATHTR---ALACIHEEGSPSPR 270
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1111 Y-LEKLTISIVINGTQGAEEVLRHAHEQLKEAQAVPATLPELEATKASLKRLAQAEAQ 1169
QY      271 TPTKSAIHLKEDIKLVGSLP--RKYSESMDLSLLSPQTKYALQSGMPNINEMKIRSIB- 328
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1170 QPTFDALRDELGAQEVGERLQORHGERDLEVRWRERVAQVLERWQAVLAQTDLQRLEL 1229
QY      329 -----QNRYYQF-----QQTIQKAKOLNSSSSSEASLAGQQ---EPLFNHS 366
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1230 EQLGKRVRYRESADPLGAWLQDARRRQFIQAMP LADSQAQVREQLRQEQALLERIEHG 1289
QY      367 EILAOSSRP-----NLHEVGRKSVKSGPVKNPKHR-----RS 398
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1290 EKVEECQFAKQIYNAIKDYELQLVTYKA-QLEPVASPAKPKVQSGSESVIQEYVDLRT 1348
QY      399 HELSPSDAPS-----NESLYENETVAISPPKKQVQEDTTLPRSAAPANASARSSAHAW 452
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1349 HYSELTTTTSQVIKIFSETLRMEEBERLAEQORABERERLAEVEALEKORQLAEHAHQ 1408
QY      453 PHAQSKFKFLAQTMSLMKKPATPRKVRDTISQPSVKLYDSEL--YMQTCINPPPPAATTI 511
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1409 AKQAQER-EAKELQRMQEEVVRREAAVDAQOQRKSTQEELQQLRQSEASIQAKARQA 1467
QY      512 DPFLASTWYLD-----QAVDRHQADFKNWLNALVSPADLADANN-KIDVGKLF 561
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1468 EAAERSRLRIEETIRVVRVRLQLEATERQGGAGELQALRARAEEAABAQROQAEARL 1527
QY      562 NEV-----RNKELVVAPTKESQSNVYLTKYRLET-----LRKAAVELF 599
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1528 RQVQDESQRKQAEVELASRYKAETEAAREKQALQALEELRLQAEAEFRRLQAEVERA 1587
QY      600 FSEQWRLPC-----SKVAVVYNKQALRIRSDRNHLHDVVMQRTILELLCFNPLW 649
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1588 RQVQVQALTAQSAEAELOSKASFAEKTAQLERSLQEEHVAVAQLRBEAE-----1638
QY      650 LRLGLEVVFGKIQMQSNRDIVGLSTFIL--NRLPRNKCEBQRYSKAYTLTVEYAEITKK 707
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1639 -RRACQQAABARAEABRE---LERWQLKANEALRLRLQAEVVAQKSLAQABAE---1690
QY      708 HSLQKILPLLPLDQAKOKRIVKINPCLFVKKS PHKETKDILLRPSSELLANIGDITREL 767
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1691 -----KQKE-----BAAREA 1700
QY      768 RRLGVYLQH--ROTFLDEFYAFNNLAVDLRDGVRLTRVVEVILIR-----DDLTRQLRV 820
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1701 RRRGKAEQAVRQRELAEQELEKQRLAEGTAQQLAAEQELIRLRAETGEQEQORQLLE 1760
QY      821 PAISRLQIRFNVKIALGALGANFOLGDDIAAQDIVDGHREKTLISLLWQLIYKFRSPKFH 880
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1761 BELARLQF-----EAAAAATOKRQELEELA-----KVRAE---1790
QY      881 AAATVLQKWRERHVLHVVIQRRIRSHKELMRHRAATVIOAVFRGHQMKYVKLFTERTQ 940
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1791 -----MEVILLASKARABESRS-----TSEKSK 1813
QY      941 AAILQKTRRYLAQKQLYQSYHSIITIQRWRAQQLGRQ-HRORFVELREAAIFLQRIW 999
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1814 QRLAEAGRFELAEAAA-----RRLALAEAKRQR--QLAEEDAAQRRAE 1857
QY      1000 RRLFAKULLAAAFARLQSRQKQAAAASYIQMOWRTYQLGRIQRHBFRLQORDLIMFYQR 1059
DQ      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ      1858 AERVLAEKLAAGIAGATRL---KTEAEIALKEAEENERLRLAEDEAFQRRRL-----EE 1909
QY      1060 RMRSKWSMLEQRKEFQQLKAAINIQQRWAKL-----SMRKNADYLAURSSVLKQVAY 1114

```

1910 QAAQHKADIER--LAQLRKASDSELRKQGLVEDTLRQRRQVEEITALKASFKEAAA- 1966  
1115 RKATQMRIDRNHYSLRKNV-ICLQORLRATMKWEQR-----ENVILRNASTLVOKRY 1169  
1967 GKAELEBLGR-----IRSNABDTLRKSKQAELEAARQQLAAAEERREAESEVOKSL 2021  
1170 RMRQMIQDRNAYL--RTRKCIINVRWRATLQMRERKNVHLHTQTTTKRIQIKFRAK 1226  
2022 AAEFEAARQKAALEVEERLKNVFEARLRERAEQESARQLQAEAAQKRLQAEKHAH 2081  
1227 REMKQRAEFLLQKVTVLVQKRRRALLQMRKERQBYLHLREVITIKLORFHAQKSWRFM 1286  
2082 AFAYQCKQELQ-----QTLQSQSVLDLQGEAE-----AARAAAEAEAEARV 2125  
1287 RAKYGTGAOAVSCLQMWNRHLLRKEERNSFLQIRQAAITLQRRYRPARLNMIKLSYAO 1346  
2126 QAEREAAQA-----RRQVEAEERLKQSA-----EEQAQARAQ 2157  
1347 LKQAAITIQ-----TRYRAKKAMQKVLYOKO-REAIKVVORRY-RGNLEMRKQLEVTYQKQ 1401  
2158 AQAAEKLRKEAEQEAARRAQAEQAAIRQQAADAEKHKKFAEQTLRQKAQVE---Q 2213  
1402 RQAVIRLQKWSIRDMRLCKAGYRRIRLSSLSIQKWRATVQARRQ-----EIFLSTI 1456  
2214 ELTTLRIQ-----LEETHQK-NLLDEELQRLKAEATAARQSQSQVEEELFSVRV 2262  
1457 RKVRL--MQAFTRA-----TLIMRQ-----ORREFEMKRAAVVIQRRFARCAMLKARQDYQ 1507  
2263 QMEELSKLKARIEAENRALITLKDNTQRYLOEAEAKQKQVAEBAARLSVAQAEARLRQ 2322  
1508 LIQSSVILQVRFRANRSMK-----QARQFVQLRTIAVHLOQ-----KFRGKRLMIQKNC 1559  
2323 LAEED--LAQQRALAEKMLKEMQAVQEAIRLKAEBELLQOKELAQEQARRLOEDKEQM 2380  
1560 FOLLRCMSPGFQARAGFMARKFOALMTPF-----MMDLIR-----QKRA 1600  
2381 AQLABETQGFQ---RTLAEARQRLQMSAEABERLKLVAEMSAQAARAEEDAQRFRQA 2437  
1601 AKVIQRYWGYL-----TPRQKHQGLLDIRKRIQRLQAKAVNSRVCKVO 1647  
2438 EBTGKHLRTELATQEKVTLVQTLEIQOQSDHDAERLREAIAREKE-----KIQ 2490  
1648 EAVRFLGRFLASDALAVLSQDLRLST 1675  
2491 QEAKLLQ---LKSEEMQTVOEQELQET 2515  
  
RESULT 10  
A39638  
plectin - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: A39638; S21876  
R:Wiche, G.; Becker, B.; Luber, K.; Weitzer, G.; Castanon, M.J.; Hauptmann, R.; Stratawa  
J. Cell Biol. 114, 83-99, 1991  
A:Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with  
A:Reference number: A39638; MUID:91268156; PMID:2050743  
A:Accession: A39638  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-4687 <WIC>  
A:Cross-references: EMBL:X59601; NID:91292885; PIDN:CAA42169.1; PID:91561642  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S  
C:Keywords: cytoskeleton; transmembrane protein  
F:6-103/Domain: ribosomal protein S10 homology <RS10>  
F:184-399/Domain: alpha-actinin actin-binding domain homology <ACT>  
  
Query Match 3.2%; Score 301; DB 1; Length 4687;  
Best Local Similarity 19.0%; Pred. No. 6.4e-08;  
Matches 347; Conservative 284; Mismatches 686; Indels 512; Gaps 76;  
QV 52 KTVLQKSPGTAGKWTWS---VVSAAVQOKRMSAAAPPSKQTVRTVAPSRPAWAHPPP 108

Dd		1189	KYLAPEPSAPPTURSELELTKGLEQVRSALYLEKLTISLVIRTOGA-----1241
Qy		109	QAPLVEKNVYKTQBEPVVISPPQRSKLKENLSMPTFGNLLDV-----IDMLRFTPLTETRG164
Dd		1242	-----EVLKTHEHLKEAQAVPATIQE-----LEVTKASLKLR-----AQA1279
Qy		165	KGOATVPFDNLAAWPTPTLKGNVKSCANDMRPRITPDDBEQPATNKTFDVKHSETINI224
Dd		1280	EAOQPVF-----NTLR-----DELGAQEVGELQOORHGHR---1310
Qy		225	SLDITDCSRIDGPHTPLNKTTIVHATHTRALACIHEBEGSPRPPTPKSAIHDLUKRDI284
Dd		1311	---DYEVERWRERVTLERWQAVIAQTDIR-----OREL1342
Qy		285	KLVGSPFKRYSBSMKDLSLLSPQTKYAIQSGMPNLNMKIRSIEQNRYQEOQOIQIKAD344
Dd		1343	EOLGRQLAYRESADPLSLWLQAK-----SRQEIOQAVP1377
Qy		345	LNSSSSSRASLAGQQ---EFLFNHSEILAOSSRFNLHVGRSKV-----386
Dd		1378	IANSQAARBOLRQEKALLFEIERHGEKBECOKFAKYI--NAIKDYBELQITYKQALEP1435
Qy		387	-GSPVKNPKHRSHELSPSDAFSNESLYXNETVAISP-----PKQR427
Dd		1436	VASPAPKKPVQSGSSSVIOEYVDLTRYSELTTLTTSQYIKFISETLRMBEEERLEAQOR1495
Qy		428	VEDTTLPRSAAPANASARSSAHAMPHAOSK---KFLAQTWLSLMKPAPTFRKVDRTSIQ484
Dd		1496	ABER---ERLAEVEALEKQRLAEHAHQAKAQABELEARELQRRMEEVTRREEAAVDAQ1552
Qy		485	PSVKLYDSEL-YMQTCINPDPAATTIIDPFLLASTMYLDB-----QAVDRHQADF534
Dd		1553	OQKRSTIOEBLOHLROSSEAEOAKAQOEBAABERSRMRIEBEIRVVRLOLETTERRQGAE1612
Qy		535	KMINALVSI PADLDADLNN-KIDVGLKFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRK593
Dd		1613	DELOALRAPAEBAEAQKRAQOEAEELRRQVODES-----QKRQ1652
Qy		594	AAVELFPFBQMRLPCSKVAVYNKQALRISRDNHLHDVVVMQRTILELLC----FNPLW649
Dd		1653	AEAEL---ALRVKAEBAAEREKQALQALDELKL-----QABEAERWLCOAEARARQ1702
Qy		650	LRLGLEVV-FGEKIOMQNRDIVGLSTFILNRL-----FRNKCEE-QRYSKAYT696
Dd		1703	VOVALETAQKSABEVLQSKRPSFAEKTAQLERTLQBHEVTVTOLREBAERRAQQAABAER1762
Qy		697	LTEYEATTKHSJL---OXKLFLLPFLDOAKORIYKHNPCLFVKKSPHETKDILLRFS753
Dd		1763	AREBAERELRWOLKANEALRLUQAEVAAQQSILAQ-----ADAEBQXE-----1807
Qy		754	SELLANIGDITRELRLGYVOLR--ROTFFDEFDYAFNNLAVDLRDGVRLTVVVEVILRL811
Dd		1808	-----EABRRARRRKAEQOAVRQRELABQELEKQORTGTGAQORLAABEQELIRLR1859
Qy		812	DDL-----TQLRVPAISRLO-----RIENVKLALGALGEANFOLGGDTAODIV856
Dd		1860	AETEQGEHORQLLEEELARILQHEATAATQKROBLEABELA-KVRAEMEVLLASKARABEES1918
Qy		857	DGREKTLISLLWOLIYKFRSPKPHAATVLOKWRRHLHWVVIQRRIRHKELMERHRAAT916
Dd		1919	RSTSSEKSKORLEABAGFRLEABEAR-----LRALAEARRHRELAEDDAARQ1967
Qy		917	VIOAVFRGHQMRVYVKLFKTER--TOAAIIQKPTR-----RYLAQKQLQSYHSIITI969
Dd		1968	RABA--DGVUTEKUALISEATRUKTAEIALKEAENERILRLADEAFQRR-----2018
Qy		970	RWRRAOQLGRQHR-----QRVFEIREAA-----IFLOQIWRRRLFAKKLLAAETA7101
Dd		2019	--RLBEOAAQHAKDIBERLAQLRKASESELERQKGLVEDTLTORRQVEEBEIMAL-----207
Qy		1017	LORSQKOQAAAASYOMQWRTYOLGRIO-----RHEFLRQSDLIM-----105

Db 2071 --KASFEAAAGKABLE---LELGRIRSNADMTMRSELAEQEAARQORAAAEERQRRE 2125  
 QY 1056 ---FYQWRMRKSWMLBOKK---BFOOLKRAINIQQWRRAKLSMRKNADYALRSV 1108  
 Db 2196 ABERVQSLAEBEAAARQKVALEVERLK-AKVBEARKLRERABEGBAQOLQAQEAQ 2184  
 QY 1109 LKQVAYRKATIOMRIDRHHYSLRKNVTCLOQRRAIMKMRQRENYL-RLRN----- 1160  
 Db 2185 KRLQAEEKA-----HAFVQQRBEELQOTLQEQNMLERLSEAEAAAR 2228  
 QY 1161 -ASILVQRYMRQOMIQDRNAYLTRKCIINQWRWRATLQ-----MRERKNYLHL 1212  
 Db 2299 AABEABEAREQAEARQAQSRKQVEABRLKQSAEBQQAQAQAQAABEKLREAB----- 2283  
 QY 1213 QTTTKRIQIKFRAKEMKQRAEFLQKVTLLVQKRRALLQMKERQEXYHLREVTIK 1272  
 Db 2284 QEAARAAQABQALQKQAADEMKHKFA-----BOTLRQAQVQBELTTLR---LQ 2334  
 QY 1273 LQRRPHAQKSM---RFMRKRYGTQAAVSLQ-----MEWRNHL 1308  
 Db 2335 LEETHD-QKSILDEBLQRLKAEVTEAARQSRQVEBELFSVRVQMBELGKLKARIEAENRA 2393  
 QY 1309 LKREKENS---FLQURQAAITLQRRYARLNMIKQIKSYAQLKQAA---ITQTRYRAKKA 1363  
 Db 2394 LILHDKDNTQRFLEEBEABKQKQVABEAAARLSVAAQ--EAARLQALABEDLAQORALABKM 2451  
 QY 1364 MOKQVLYQ-----KQBAILIKVQRRYRGNLEMRKQIEVYQKQRAVIRLOKWRBSI--- 1415  
 Db 2452 LKEMQAVQEAATRLKAEABELQOQKELAQEQARRILOADKQWAAQOLVEETQGFORTLEAE 2511  
 QY 1416 --RMRLCKAGVR-RIRLSSISIQKWRATVOARFORE-----IFLSTIRKVL 1461  
 Db 2512 RQROLMSAEARLKLRAEWS-RAQAAEADAQRFRKQAEIEGKLRHTELATQEKVTL 2570  
 QY 1462 MQAFIRATL-LMRQOR-REFEMKRAAAVQIR-----RFRARCAMLKARDYOLLIQSSVI 1514  
 Db 2571 VQ-----TLETQROQSDQARLEAREIAELEREKELQKQAKLLQKS--EEMQTVQOQOI 2624  
 QY 1515 L-----VORKFRANRSMQARQEFVOLRTIAVHLOKF-----RGKRLMIQORNCQFLLR 1564  
 Db 2625 LQETQALQKSFSEKDSLLQERFIQEB--KAKLEQLFQDEVAKAKQIQEIQ----- 2674  
 QY 1565 CSMPGFQARAGFMARKEFQALMTPEMDLI-----ROKRAKVQIRYWRGY--LIR 1614  
 Db 2675 -----ORQOQMEQEKQELVASMEAEARRQREAEAGVRRKQBEELQRLQEQ 2718  
 QY 1615 PROKHQGLL-----DIRKRIALQREAKA 1638  
 Db 2719 QROQOEXLLAEENQRLRERLQRLBEEHRA 2747  
 RESULT 11  
 T05200  
 myosin heavy chain F4110.130 - Arabidopsis thaliana  
 N;Alternate names: protein F4110.130  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Dec-1999  
 C;Accession: T05200  
 R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hoheisel, J.; Mew  
 submitted to the Protein Sequence Database, February 1999  
 A;Reference number: 215402  
 A;Accession: T05200  
 A;Molecule type: DNA  
 A;Residues: 1-1375 <BRV>  
 A;Cross-references: EMBL:AL035525  
 A;Experimental source: cultivar Columbia; BAC clone F4110  
 C;Genetics:  
 A;Map position: 4  
 A;Introns: 48/3; 90/3; 139/2; 191/3; 211/2; 263/3; 313/3; 359/2; 408/2; 442/2; 461/3; 49  
 ; 1134/3; 1155/3; 1214/3; 1238/2; 1271/3; 1298/3; 1326/2  
 A;Note: F4110.130  
 C;Superfamily: myosin MYO2; myosin motor domain homology  
 C;Keywords: P-loop

F;63-714/Domain: myosin motor domain homology <MMOT>

Query Match 3.2%; Score 300.5; DB 2; Length 1375;  
 Best Local Similarity 21.5%; Pred. No. 1.2e-08;  
 Matches 263; Conservative 178; Mismatches 406; Indels 379; Gaps 65;  
 QY 320 NEMKRIEONRYOEOQIQIKAKKDLNSSSSSEASLAGOQEFLEFNHSEIL-AQSSEFNHL 378  
 Db 201 NAKTVRNDSNRFGKVEIQF---DTNGRISG---AAIRTYLLERSRVRRITDEPRTVH 253  
 QY 379 EYGRKSVKGS-----PVKNPHKRRSHSLSFS-----DAPSNESLYRN-----ETVAISPP 423  
 Db 254 CFYQLCASGDAEKYKLSNP--RQPHYLNQSKTYELEGVSSAAEYKNTREANDIVGISQD 311  
 QY 424 KQORVEDTTLPRSAAPANASARSSAHAWPRAQSKFKLAQTMSLMKPKPATPRKVRDTSI 483  
 Db 312 BOEGT-FETLAAIHLGNVEFSSGREH-----DSSVVKDPESRH-----HL 351  
 QY 484 OPSVKLY--DSELVMOT-CINP-----DPFAATTTIDPPLASTMYLDEQAVDR 528  
 Db 352 QMAADLFKCDANLLASICTRSILITREGIIIKALDPNAAVISRD-TAKTVY----- 402  
 QY 529 HQADPKMNLALVSPADLDADLNKKIDVGKL-----FNEVRNKE-----LVVAPTKEQS 579  
 Db 403 --AHLFDWL--VDKINKSVQDPESEFQIGVLDIYGFECFKNSPEQFCINPANEKLOQH 458  
 QY 580 MN-VLTQVLETLKAAVELFSEQMRPLPCSKVAVYVKNQALRIRSDNLHLDDVVMQRTI 638  
 Db 459 FNEHVFKMEQDEYRKBEINWSYIE-----FIDNQDV-----LDLIEKPI 498  
 QY 639 LELLIL-----CFNPLWLRLGLEVVFGCKIQMOSNRDIVGLSTFILNRLFRNKCEBORYSKA 694  
 Db 439 GVIALLDKACMP-----RSTHESFSMKLF-QNFRPHRLEKFKFSET 540  
 QY 695 YTLFEYEAETIKHSLQKILFLPLDOAKQRIYKH-----NPLFV-----KKSPHKE 744  
 Db 541 DFTLSHYAGK-----ATFLDKNRDVTIVEHCNLLSSSKCPFVAGIFPSAPEES 588  
 QY 745 TKDILLRPS-----ELLANIGDITRELRGLVYLOHROTFLDEFYAFNNLAV--DL 795  
 Db 589 TRSSSYFSSVSSRFKQOLQALMETLSK--TEPHYVRCVKPNSLNR-PQKPESLSVLHQL 645  
 QY 796 RDGVRLTRVVEVILLRDDLTRQ-----LRVPAL--SRQRIENVKLALGALGEA 842  
 Db 646 RCGGVLEAV--RISLAGYPTERNYSDFVDRGLLAPEFMDENDEQALTEKI--LSKILG 702  
 QY 843 NFQLG-----GDIAAQDIVDGHREKTLISLLQWLIYKFRSPKFAAATVLOKWRHRH 894  
 Db 703 NYQLGRTKVFLRAGQIG--ILDSRRAEVLD-----ASAR----- 734  
 QY 895 LHVVIQRRIR-----HKELMRHRAATVIOAVFRGHQMRKYVKLFKTER-TQAAIILQKFT 949  
 Db 735 ---LIQRRLRFTVTHQNFISARASAIQYACRGCLSR---NAYATRNAAAAALVQKHV 788  
 QY 950 RYLAQQLQYQSYHSIITIQRWAAQGLQROHROKRFVELREAAIFLQRIWERRLFAKLL 1009  
 Db 789 RRLWS-----RCAPVKLVSAALVLOGCIR----- 812  
 QY 1010 AAATFARLORS-QKQAAAASYIQMWRTYQLGRIQRHEFLQRDLIMFVQRMRSKWSML 1068  
 Db 813 --ADSTRKFTSHQEHRAASLIQAHWHIKF---RSAFRHQSSIIAIQCRWQKLA-- 864  
 QY 1069 EORKEFOOLKRAA-----INIQOR-----WRAKLSMRKNADYLA----- 1103  
 Db 865 --KEEPRKLQVANEAGALRLAKTKLEKRLDLEWRLQLEKRLTSGEAKSSEISKLOK 922  
 QY 1104 -LRSSVLKQVAYRKATQIMRIDRHHYSLRKNVTCLOQRRAIMK-----WEEQ 1151  
 Db 923 TLESFSLKDAARLNTIN-----ECNKNAV-LEKQIDISMKEKSAVERELNGMVEL 972  
 QY 1152 RENYLRLNASILVQRYMRQOMIQDRNAYLTRKCIINQWRWRATLQWRRE-RKNYL 1210  
 Db 973 KCONALLKNSMNSLEKKNRVLE-----KELLNAKTCNNVTLOKLEAEKRC 1019



QY 1211 HLOTTTKRQIKFRAKREMKQABFLQKKVT-----LVQKERRALLQWRKEROY 1263  
 Db 1020 ELQTSVQSLEEL-----SHLENEQVIMQKTLTSPERICQILGEKHSSAVPAQNDRSV 1076  
 QY 1264 LHLREVTIKLRRFHAQKSMRFRKRGTOAAVSCLOMHR-----N 1306  
 Db 1077 FENYELLSRC-----IKENLGNDDKPLAACVYKCL-LHWRAFESSTAFNIIIEGIN 1130  
 QY 1307 HLLRKRR-NSFLQRLQAAITQRRYRARNLMIKQLKSYAQLKQAAITQTRYRAKAMQ 1365  
 Db 1131 EALKRNLRSNLFN-----ASAQRSGRAAYGVKSPFKLHGP-DGGAHIEARYPA----- 1179  
 QY 1366 KQVLYQKQREAIL-KVQRRYGNLB 1390  
 Db 1180 ---LLFQKQLTACVEXIYGLIRDNLK 1202

RESULT 12  
 A59254  
 myosin heavy chain 12, splice form 1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Sep-2000  
 C:Accession: A59254  
 R:Neurens, B.  
 submitted to GenBank, September 1996  
 A:Description: The complete cDNA for human myosin heavy chain 12, a class V myosin.  
 A:Reference number: A59254  
 A:Accession: A59254  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1855 <MEU>  
 A:Cross-references: GB:Y07759; NID:gl770740; PIDN:CAA69035.1; PID:gl770741  
 A:Experimental source: isolate 3238; dev stage adult; tissue type skin; cell type fibroblast  
 C:Comment: For an alternate splice form of the same gene see PIR:B59254.  
 C:Genetics:  
 A:Gene: MYH12  
 A:Map position: 15q21  
 A:Superfamily: myosin MYO2; myosin motor domain homology  
 F:72-751/Domain: myosin motor domain homology <MMO>

Query Match 3.2%; Score 300.5; DB 2; Length 1855;  
 Best Local Similarity 19.2%; Pred. No. 1.9e-08;  
 Matches 289; Conservative 224; Mismatches 545; Indels 450; Gaps 64;

QY 559 KLFNEVRN--KELVAPTKEEQSMNYLTKYRLTLKAAVELFFSEQMRLPCSKVAVYN 616  
 Db 410 KLFNWIQVNVNQALHSAVKQHSPIGVLDIYGFETFEINSFEQF-----CIN---YAN 458  
 QY 617 KQALRIRSDNLHL-----DVVMQRTILELL-----CFNPLMLRLIGLEVVFGBKIQM 664  
 Db 459 E---KLOQQRNMHVFKLEQEEYMKQIPLWTLIDFYDNQPCINLIESKLGILDLDECKM 515  
 QY 665 QSNRDIVGLSTFILNLFNKC---BEQVYS-KAY-----TLT 698  
 Db 516 PKGTDDTWAQKLYNTHL--NKCALFEKPRLSNKAFTIHPADKVEYQCEGFLEKKNKDTVF 573  
 QY 699 EYAEYTIKHSLOKLIPLPFDQAKKRI-----VKHNPCLFVKKSP----- 741  
 Db 574 EQIKVLKSSKFK-----MLPELFQDDEKALSPTSATSSGTPITRTPAKTKRPGQMAK 629  
 QY 742 -HKETDKILLRFSSELLANIGDITRELRRLGYVLQHRQTFELDFDYAFNNLAVDLRDGVR 800  
 Db 630 EHKKTVGHQFNSLHLL-----METLNATPHVVRCKIKENDFKL-PFTFDEKRAVQ 679  
 QY 801 LTRVVEVILLRDDLTQLRVPD-----ISLQRIQFNVKLAG----- 837  
 Db 680 QLRACGYL-----ETIRISAAGFFSERWTYQBFESFYVLMQKQVLSDRKQTCRNMLE 732  
 QY 838 --ALGEANFOLGGDIAAQDIVDGHREKTLISLLQWLIY--KFRSPKPHAAATVLQKWRRH 893  
 Db 733 KLILDKDKYQFG-----KTKIFPRAGQVAYLEKLRAKLRACIRIQK-TIRG 779

QY 894 WLHVVIQRIIRHKELMRHRAATVIOAVERGHOMRYKVKLFKTERTOAAIILOKFTRRYL 953  
 Db 780 ML-----LRKKYLEMEKAAITMQRYVYRGYQARCYAKFLR--RTKAAITIQKYMWMV 829  
 QY 954 AQKQ-----LVQSY-----HSIITIQ---RWVRAQQLGQHQRF 985  
 Db 830 VERRYKIRAAATVLOSYLGRFLARNRYKILGRHKAVTIQKRVGW-----LARTHYKKS 885  
 QY 986 VELRRAAIFLQRIWRRRLFAKLLAAAEATARLQRSQKQAAASYIQMWRTYQLGRIQRH 1045  
 Db 886 M---HAIITQCCFRMAKRELKKLKEAR-----SVERY 918  
 QY 1046 EFLR--QRDLIMFQVRMRSKWMLERKEFOOLKAAAINIOORWPAKLSMRKCNADYLA 1103  
 Db 919 KKLIRIGMENKIMQLQKVD-----EONKDYKCLVEKLTNLEGINSETKLSRDLERLQ 972  
 QY 1104 LRSSVLKVOAYRYKATIQMRID--RHYYSLRKNVICLOORLRAIMKMRQRENLYLRNA 1161  
 Db 973 LSEEEAKVATGRVLSLOEBIAKLKDLQTRSEKKCIBEHADRYKQETQLVSNLKEEN- 1031  
 QY 1162 SILVQKRYMRQOMIQDRNAVLRTRKCIINVORRWRAITLQMRERKNYLHLQTTTKRIQI 1221  
 Db 1032 TLLKQKEALNHRIVQAAKEMTETMEKCL-VBETKQLELDLNDERLYQNLLNEFSLE 1090  
 QY 1222 KFR-----AKREMKKQRAAFLOLKKVT- 1243  
 Db 1091 RYDDLKEEMTLMVHVPKPGHKRTDSTHSSNESSEYIFSSBIAEMEDIPSTEEFSEKKVPL 1150  
 QY 1244 ---LVVQKRRRALLQMKERQEVYHLHREVITIKLQRRFHAQKSMRFAKVRGTQAAVS 1298  
 Db 1151 DMSLFLKQKRVTELEQEKVQMDLDRKEBQVL-----RSKAKEEERPOINGAELEYE 1204  
 QY 1299 CLQMWNRNL--LRKERNRSFQLRQAAITLQRR-----YRARLNMIKOLKSY 1344  
 Db 1205 SLK---RQLESENKKLKNELNELKA---LSEKSAPEVTAPGAPAYRVLMQLTQSVSE 1258  
 QY 1345 AOLKQAAITQTRYAKAKAMQKVVL-----YQKOR-----EALIKV 1381  
 Db 1259 LDVRKEEVL---LRQLSVSQKEAIQPKDDKNTMTDSTILLEDDVQKMKQGEIAQAYIGL 1315  
 QY 1382 QRRYRGNL---EMRKQIE---VYQKQRA---VIRLOKQWESIRD-----MRLC 1421  
 Db 1316 KETNRSSALDYHELNEDEGELWLVEGLKQANRLLESQLOKSKSHENEAALRGETQSLK 1375  
 QY 1422 KAGYRIRLSSLSIQKRWATVOARRQRIIFLSTIRKVRIMQAFIRATILIMRQOREFEM 1481  
 Db 1376 EENNRQQLLAQNLQLPPEARIEASIQHEITRLTNENLDLMEQ-----LEKQDKTVRKL 1429  
 QY 1482 KRRAAVIQRFRFARCAMLKARODYQIIOSSV--ILVQRKFRANRSMKQARQEFVQLRTI 1539  
 Db 1430 KQQLKVFAKKIGLEVGQEMENISPGQIIDEPIRPVNIIPRKEKDFQGMLEYKKKDEQKLVK 1489  
 QY 1540 AVHLQOKFGRKLMIBQRNCFQLLRCSMPGFQARARGFAR-----KRFOALMTPEM 1591  
 Db 1490 NLILELRPGVAVNL-----IPGLPAYILFMCVHRHADYLNDDQKVSLSLTSTI 1537  
 QY 1592 MDLIROKRAAKTQOR-----YWRGYLIR-----RQKHQGLL 1623  
 Db 1538 NSI-----KKVLKRGDDPFTVSWFLSNTCRFLHCLQKYSGEFGFMKHTSONEHCLT 1591  
 QY 1624 DIPKRIAQLRQEKAVNSRVCKV-QEAVRPLRG---RFTASDAL----- 1663  
 Db 1592 NF--DLAEYRQ---VLSDLAIOIYQOQLVRVLENIQFMIVSGMLHEHTIQGVSGVKPTGL 1646  
 QY 1664 --AVLSQDLRSTVPHLLMWCSEPMSTFCYIGMAQAIRSEVDKQILERSRIIAN--- 1717  
 Db 1647 RKTSSIADEGTTVLSILRQLNSFHSVMCHGMDPELIKQVVKQMFYIIGAITLNNLL 1706  
 QY 1718 ---LARYNSTTV-----NTPQEGG-----LVTTIAQML--LRWCDDKSHI 1751  
 Db 1707 RKMCSKSGKQIRYNVNSQLEEWLRDKNLNMNSGAKETLEPLIOAAQLQVKKTKDDDAEA 1766  
 QY 1752 FNTILCTLI 1759







Qy	558	GKLFNEVRNKL	VVAPTKEBQSMNYLTKYRLETIRKAAVELFFSEQR	-----LPC	608
		:	:	:	:
Db	478	EAFI	--OLKEVAKERKELEEL-YLVKKQKQDKENEL-LPFKEQLKHOADFENELEA	532	
Qy	609	SKYAVYVVKOALIRS	-----DRNLHLDDVMQRTIILELLCFNPLMLRLEGVVFEKIQ	663	
		:	:	:	:
Db	533	KQOELFEAKHLE	-RSFIKLEDEKOLNTRAOQ-----IANEFSQ	571	
Qy	664	MOQNRI	VGLSTILNRLFN-KCEEORYSKAVLTTEEVAETIKKHSLOKILFLLPFLDQ	722	
		:	:	:	:
Db	572	LKTDKSKSAD	FELMLQNEYENLOEKQKLFQERTYFERNAAVLNRLQKRELL	-----	826
Qy	723	AKOKRIVKHN	PCILFVKVSPHETKDILLR-FSSELLAN-----IGDITRELRRLGYVL	774	
		:	:	:	:
Db	627	-----	QOKETLQDQTKSFQERLINQREHKELVASVEKQEKILKKL	668	
Qy	775	Q-HROTFLDE	FAYFNLA--VDLRDGVELTRVVEVILLRDDTLQLRVPALSRQIRF	830	
		:	:	:	:
Db	669	QDFSQISLN	-----ASKNLAERMAIPEKIEIBATEKQLLND-----VN	708	
Qy	831	NVKLALGAL	GEANFOLGGD-----IAAQDITVDGHRKTIISLLMQLIYKFRSPKFHAAATV	885	
		:	:	:	:
Db	709	NAEVIQAD	LAQLNQLNQERSELQNAKQRIADFHNDLSKKL-----	749	
Qy	886	LQKWRRHVL	HVIOQRIRKHELM-----RRHRAATVIOAVFRGHMRKVVYKVLPKTERTQA	941	
		:	:	:	:
Db	750	-----	NEVELSLOKRELOLEANQOKHOSYN--QAYFEG-----ELDKLNREKQ	793	
Qy	942	AI--ILQKTR	RYVLAOKOLYOSVHSIITIQRWRAQOLGRHQRFVVELFEAAIFLQRIW	999	
		:	:	:	:
Db	794	AFNLNRKQT	MEVDALQRLSDKHQALNMQO-----BELDKTH-----ELNNA--FLNHDA	843	
Qy	1000	RRBLFAK	LAAAEATAR--LORS--QKQAAASYI--QMOQ-----RTYQLGRI	1042	
		:	:	:	:
Db	844	DOKSLQD	QATVKTOKLIDLERSALLEKQREFAENVAGFRHWSNKTSLQKXIYELTKK	903	
Qy	1043	ORHEFT	RQDLINFVQRRMRKWSMLBORK--FFQOLKRAINIQRWRAKLSMRKNAD	1100	
		:	:	:	:
Db	904	QESQEQK	ETELAFSDLOKDYQVPELOKQDFRQ-----IBAKQRELDKLA-EKKNQV	957	
Qy	1101	YLALRSVL	KVQAYRKATIOMRID-RNHYYSLRKNVICLOQLRLAIMKREQ-----	1151	
		:	:	:	:
Db	958	KLELDN	FPALQNKQDVTQAQLELEREHQJNLQETAFAQNANESLLKQREQLTKKIQAF	1017	
Qy	1152	-----	RENYILRLRNASILVKRYMRQOMIORDNAVLRKCIINQRRWRATLOMRE	1205	
		:	:	:	:
Db	1018	HYELKKRN	QFLALKGKRLFAKEQDQOKD--QBIN-----WR-----FKQF	1056	
Qy	1206	RKNVHL	QITTKR-----IQIFKAKREMKQBAEFLQKKVTLVVOKERRAL	1253	
		:	:	:	:
Db	1057	EKEYTP	DEAKKRELELEKIRSLQSNVLEBRKREKATDTFNLNKVOHNTQINRD--	1114	
Qy	1254	LQMKERQ	EYHLREVTIKLQRFHQAOKSMRFW-----AKYRGTAQAVSCLQMEHRN	1306	
		:	:	:	:
Db	1115	-QLNSQIR	QELLERKNQFORSNEANAKKAFILKELRSFASNKLQKEALAIQKLEPDKD	1173	
Qy	1307	HLLRKERN	SFLQRAITLQRRYRARLNNIOLKSYAOLKQAATIQTPYR-----AKK	1362	
		:	:	:	:
Db	1174	EQQKKE	LQATLQLEQPFKEQ-----NFDIEKQ-----RQLVAIKTQCEKLSDEKK	1220	
Qy	1363	AMOKQV	LYLQKREAI-----KVQRYVRGNLEMRKQIE-----VYQKQCAVI	1406	
		:	:	:	:
Db	1221	ALMQVL	VELKNLSQTYLANKNKABYSQQQLQCKYTNLLDKENLERTKQOLDKKHSIFA	1280	
Qy	1407	RLOKWR	SIR-----DMRLCKAGYRIRLSLSLSTQKWRATV-----	1443	
		:	:	:	:
Db	1281	RLTKFAND	LAFEXKQLLKAQRIVDDKNRLKENERNLHFLSNETERK-RAVEQOISYPE	1339	
Qy	1444	-QAPRQREI	FLSTIRKVR-----LMOAFIRATLL-----MQORREFEMKQRAAV	1487	
		:	:	:	:
Db	1340	KORKQA	YDAILASHKEYKKEGELQKILVELETKTKLIANNDFAKFSQREFFNORIKLIL	1399	

Search completed: September 13, 2004, 17:43:44  
Job time : 51 secs

*This Page Blank (usp1c)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:36:33 ; Search time 23 Seconds  
(without alignments)  
4213.155 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVSPVLEVACKETQLLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	324	3.4	1407	1 TRHY_RABIT
2	323.5	3.4	1549	1 TRHY_SHEEP
3	317.5	3.3	1829	1 MYSA_CHICK
4	303.5	3.2	2779	1 LVA_DROME
5	301.5	3.2	1855	1 MYSA_HUMAN
6	301	3.2	4687	1 PLE1_RAT
7	295.5	3.1	1828	1 MYSA_RAT
8	294	3.1	4684	1 PLE1_HUMAN
9	287	3.0	4473	1 PLE1_CRIGR
10	284.5	3.0	1805	1 HWM2_MYCGE
11	281	3.0	1742	1 MY5C_HUMAN
12	277	2.9	1818	1 HWM2_MYCPN
13	277	2.9	1898	1 TRHY_HUMAN
14	274	2.9	1498	1 GOA3_HUMAN
15	271.5	2.9	1853	1 MY5A_MOUSE
16	270	2.8	2442	1 CEP2_HUMAN
17	267.5	2.8	2663	1 CENE_HUMAN
18	260	2.7	1979	1 TRIA_HUMAN
19	260	2.7	2245	1 MGSJ_DICDI
20	258.5	2.7	2238	1 GOA4_MOUSE
21	257.5	2.7	1849	1 MY5B_HUMAN
22	254.5	2.7	1411	1 EEA1_HUMAN
23	254	2.7	3210	1 CENF_HUMAN
24	246.5	2.6	2022	1 ANTI_ONCVO
25	244	2.6	1846	1 MY5B_RAT
26	243.5	2.6	1453	1 Y373_BOVIN
27	238.5	2.5	3674	1 SPCK_HUMAN
28	234.5	2.5	1941	1 MHX2_HUMAN
29	232.5	2.4	1679	1 GCC2_MOUSE
30	230.5	2.4	1211	1 SBCC_PSEAE
31	229	2.4	1539	1 Y373_HUMAN
32	228.5	2.4	1679	1 Y109_YEAST
33	228	2.4	2611	1 BPLE_MOUSE

## ALIGNMENTS

RESULT 1  
TRHY\_RABIT  
ID TRHY\_RABIT STANDARD; PRT; 1407 AA.  
AC P37709;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Trichobyalin.  
GN THH.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fietz M.J., Rogers G.E.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Intermediate filament-associated protein that associates  
CC in regular arrays with keratin intermediate filaments (KIF) of the  
CC inner root sheath cells of the hair follicle and the granular  
CC layer of the epidermis. It later becomes cross-linked to KIF by  
CC isodipeptide bonds. It may serve as scaffold protein, together  
CC with involucrin, in the organization of the cell envelope or even  
CC anchor the cell envelope to the KIF network. It may be involved in  
CC its own calcium-dependent post-synthetic processing during terminal  
CC differentiation.  
CC -!- SUBUNIT: Homodimer (Probable).  
CC -!- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as  
CC the inner root sheath (IRS) of hair follicles and medulla, and in  
CC the filiform papillae of dorsal tongue epithelium (Probable).  
CC -!- DEVELOPMENTAL STAGE: Expressed during late differentiation of  
CC the epidermis.  
CC -!- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand  
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely  
CC alpha-helical, configured as a series of peptide repeats of  
CC varying regularity, and are thought to form a single-stranded  
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is  
CC the most regular and may bind KIF directly by ionic interactions.  
CC Domains 5 and 7 are less well organized and may induce folds in  
CC the molecule. Domain 9 contains the C-terminus, conserved among  
CC different species.  
CC -!- PTM: Substrate of transglutaminase. Some 200 arginines are  
CC probably converted to citrullines by peptidylarginine deiminase.  
CC -!- SIMILARITY: In the N-terminal section; belongs to the S-100  
CC family.  
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z19092; CAA79519.1; -.

34 227.5 2.4 3259 1 GOB1\_HUMAN O14789 homo sapien  
35 227 2.4 3678 1 DMD\_MOUSE P11531 mus musculus  
36 226.5 2.4 3214 1 BPA1\_HUMAN Q03001 homo sapien  
37 226.5 2.4 3911 1 AKA9\_HUMAN Q99996 h a-kinase  
38 225.5 2.4 5171 1 BPEA\_HUMAN O94833 homo sapien  
39 225 2.4 3685 1 DMD\_HUMAN P11532 homo sapien  
40 224 2.4 1756 1 PEPL\_HUMAN O60437 homo sapien  
41 224 2.4 2230 1 GOM4\_HUMAN Q13439 homo sapien  
42 222.5 2.3 3321 1 PCN2\_HUMAN Q95613 homo sapien  
43 222 2.3 1755 1 PEPL\_MOUSE Q92619 mus musculus  
44 221.5 2.3 2033 1 EVPL\_HUMAN Q92817 homo sapien  
45 221.5 2.3 2871 1 DESP\_HUMAN P15924 homo sapien



CC CC Comment=Additional isoforms seem to exist;  
 CC CC Names=Long;  
 CC CC IsoId=P22793-1; Sequence=Displayed;  
 CC CC Names=Short;  
 CC CC IsoId=P22793-2; Sequence=VSP\_000847, VSP\_000848;  
 CC CC TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as  
 CC CC the inner root sheath (IRS) of hair follicles and medulla, and in  
 CC CC the epithelia of the tongue, hoof and rumen.  
 CC CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand  
 CC CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely  
 CC CC alpha-helical, configured as a series of peptide repeats of  
 CC CC varying regularity, and are thought to form a single-stranded  
 CC CC alpha-helical rod stabilized by ionic interactions. Domain 6 is  
 CC CC the most regular and may bind KIF directly by ionic interactions.  
 CC CC Domains 5 and 7 are less well organized and may induce folds in  
 CC CC the molecule. Domain 9 contains the C-terminus, conserved among  
 CC CC different species.  
 CC CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are  
 CC CC probably converted to citrullines by peptidylarginine deiminase.  
 CC CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100  
 CC CC family.  
 CC CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC CC EMBL; Z18361; CAA79165.1; -;  
 CC CC EMBL; X51695; CAA35992.1; -;  
 CC CC PIR; A40691; A40691.  
 CC CC HSP; P02633; IG5.  
 CC CC InterPro; IPR001751; CaBP S100.  
 CC CC InterPro; IPR002048; EF-hand.  
 CC CC Pfam; PF00036; efhand; 1.  
 CC CC Pfam; PF01023; S\_100; 1.  
 CC CC ProDom; PD003407; CaBP S100; 1.  
 CC CC PROSITE; PS00018; EF HAND; 1.  
 CC CC PROSITE; PS00303; S100 CBP; FALSE NEG.  
 CC CC Keratinization; Calcium-binding; Repeat; Citrullination;  
 CC CC Alternative splicing.  
 CC CC  
 CC CC DOMAIN 1 91 S-100 LIKE  
 CC CC CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 CC CC DOMAIN 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 CC CC REPEAT 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.  
 CC CC REPEAT 413 448 1-1.  
 CC CC REPEAT 449 476 1-2.  
 CC CC REPEAT 477 504 1-3.  
 CC CC REPEAT 505 532 1-4.  
 CC CC REPEAT 533 560 1-5.  
 CC CC REPEAT 561 588 1-6.  
 CC CC REPEAT 589 616 1-7.  
 CC CC REPEAT 617 644 1-8.  
 CC CC REPEAT 645 678 1-9.  
 CC CC REPEAT 679 706 1-10.  
 CC CC REPEAT 707 742 1-11.  
 CC CC REPEAT 743 771 1-12.  
 CC CC REPEAT 772 796 1-13.  
 CC CC REPEAT 797 832 1-14.  
 CC CC DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.  
 CC CC REPEAT 938 961 2-1.  
 CC CC REPEAT 962 985 2-2.  
 CC CC REPEAT 986 1021 2-3.  
 CC CC REPEAT 1022 1044 2-4.  
 CC CC REPEAT 1045 1067 2-5.  
 CC CC REPEAT 1068 1090 2-6.  
 CC CC REPEAT 1091 1121 2-7.  
 CC CC REPEAT 1122 1144 2-8.  
 CC CC REPEAT 1145 1167 2-9.  
 CC CC REPEAT 1168 1197 2-10.

FT REPEAT 1198 1227  
 FT REPEAT 1228 1250  
 FT REPEAT 1251 1273  
 FT REPEAT 1274 1296  
 FT REPEAT 1297 1319  
 FT REPEAT 1320 1342  
 FT REPEAT 1343 1368  
 FT REPEAT 1369 1391  
 FT REPEAT 1392 1416  
 FT REPEAT 1417 1439  
 FT REPEAT 1440 1461  
 FT REPEAT 1462 1484  
 FT REPEAT 1485 1507  
 FT VARSPLIC 1145 1197  
 FT VARSPLIC 1251 1273  
 FT CONFLICT 1399 1399  
 FT SEQUENCE 1549 AA; 201173 MW; E72FB9FF1326E54E CRC64;  
 SQ  
 Query Match 3.4%; Score 323.5; DB 1; Length 1549;  
 Best Local Similarity 19.3%; Pred. No. 3.3e-10;  
 Matches 280; Conservative 249; Mismatches 504; Indels 421; Gaps 60;  
 QY 282 RDKLVGSPLRKYSMSK-DLSLLSPQTKYALQGMPLNE-----M 322  
 DB 308 ROKVYSPHPHQEQSSRRQEQELLERQEQQISEVQSLQEDQGRQRLKQRYDQNRW 367  
 QY 323 KIRTEQNRY-----QEQQIQIKAKDLNSSSS-----SPASLAGOQEF 361  
 DB 368 QLEESQRRRYTYAKPAQREVEEQLEKKEKQREKQREKQREKQREKQREKQREK 427  
 QY 362 LFNHSEILAQSRRFNLHEVGRKSVKGVKSPVKNPKHRSHE-----LSPSDASNSLYNE 416  
 DB 428 LQREEQLOREER-----EKRRQREKQYLEKVELWEELQOREE 468  
 QY 417 TVAISPPK-KORVEDTTLPRGAAPANASARSSAHAWPHQSKKPKLAQTMKMKPATP 475  
 DB 469 REKQREKQYLEKVEL-REEEQLOREK-----RRQREKQYLEKVELQEEELQ 521  
 QY 476 RKVRTSTQPSVKLYDSLYMOTCINPPFAATTIDPFLASTWYLDQAVDRHOADPKK 535  
 DB 522 REEREKQREKQYLEKVELQ-----EEEQLOREKRR 557  
 QY 536 WLNALVSPADLDADLNKIDV---GKLFNVRNKLVAFTKEOSNLYLTXYL---E 589  
 DB 558 -----QEREKQYLEKVELQEEQLOREKQREK-----RQREKQYLEKVELQEE 602  
 QY 590 TURKAAVELFFSEQWRLPCSKVAVYVVKQALRIISDR-----NLHLDVVMQRTILEL-- 641  
 DB 603 QLOREKQREKQYLEKVELQEEQVQVQREKQREKQREKQREKQREKQREKQREK 662  
 QY 642 ---LLCFNPLWRLGLEVVFGKIQOMQSNRDIIVGLSTFILNRLFNKCEBQRYKAYTIT 698  
 DB 663 EQQLLEEREKQREKQYLEKVELQEEQ-----LQREEREKQREKQREKQREKQREK 713  
 QY 699 EBYAETIKKHSLOKILFLPLDQAKOKRIKVNPNCLFVKKSPHKBTXDI-----LLRFS 753  
 DB 714 EELQREKQREKQYLEKVELQEEQVQVQREKQREKQREKQREKQREKQREKQREK 768  
 QY 754 S-----ELLANIGITITRELRLGYVLQHRQTFDFDYAFNNLAVDRDGVRLTRV 804  
 DB 769 REKQYLEKVELQEEQLOREKQREKQREKQREKQREKQREKQREKQREKQREKQREK 815  
 QY 805 VEVI-----LLRDDLTQQLRVPAISELQIFNVKLALGALGEANF-----OLGDIQAQ 853  
 DB 816 EQQLQREKQREKQYLEKVELQEEQLOREKQREKQREKQREKQREKQREKQREKQREK 863  
 QY 854 DIVDGHRE-----KTLSSLWQLIYKFRSPKFAAATVLQ-----KWR-----R 892  
 DB 864 RYVSKHREKQREKQYLEKVELQEEQLOREKQREKQREKQREKQREKQREKQREKQREK 918  
 QY 893 HWLVVIOORIRHKLMEHRRAATVIAVFGHQMK--YV-----KLFKTERT 939



```

Db 919 QPFAQLERQKQETTERDRKFRBEEQLLKGQREKIRYLEFEDRKFRBEEQQLRRRLRE 978
QY 940 Q-----AAIILKFTFRYLAQKLYQSYHSIIITIQWRAQQLGRQHRQR-FVE 987
Db 979 QQLRQERDRKFRBEEQLRQERDRKFRBEEQLLQ-----REQLRQERDRKFR 1027
QY 988 ----LREAIPLQRIWRRLF-----AKLLAAETARLQSQVQQAASVIMQW 1034
Db 1028 BEQLLQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQERDRKFRBEEQLLREEOQL 1087
QY 1035 RYVQLGRIGHFELQRIQDILMPVQ-----RMRKSWMLQERKEFQQLKRAAIN 1083
Db 1088 ROERNKFRBEEQLRQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQERDR 1145
QY 1084 IQORWRAK-----SMKCNAD-----YLAIRSSVLVQVVRKATIQMRIDRNHY 1129
Db 1146 KRFREEAQILKREBEEQLRQERDRKFRBEEQLLQERBEEQLLQERBEEQL 1201
QY 1130 SLRKNVICLOQLRAIMKMRQERENYVLRNASILVOKRYMRQOMIQDRNAVLRTRKI 1189
Db 1202 REEQLLQERKL-----RQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQ 1249
QY 1190 INVQRWRATLQMRBERRNYLHLOTTTKRIQIKFRAKEMKORABFLLQKVLVQKR 1249
Db 1250 --RDRKFRBEEQLRQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQ 1296
QY 1250 REALL--QMKREROYVHLREVTIKLQRRHAQ-----KSMRFRKRYRGTQAAVSLQ 1301
Db 1297 DRKFRBEEQLRQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQERDR 1345
QY 1302 MWRNHLKRRNRNSFLQRLQAATLQRRYARLNMIKQLSKYALQKAAITITRYRAK 1361
Db 1346 FHEKEHLREREQ--QLRQERLE-----GVFSQEBEQLRQERBEEQLRQERDR 1393
QY 1362 KAMQKQVLYQKQRAIKVQRRYRGNLKMKQIEVYQKQAVIRLQKWRSTDRMLC 1421
Db 1394 KFLBEEQLRQERBEEQLRQERDRKFRBEEQLLQERBEEQLLQERBEEQLRQ 1440
QY 1422 KAGYRRIRLSSLSIQKRWATVQARRQREIFLSTIRKVLQAFIRATLLMRQRRRFE 1481
Db 1441 -----LQOQYAEQFAFE-----EKRRRQEBELRQERBEEQLRQER 1476
QY 1482 KRAAVTVQRRFACAMKARQYQLQSSVLVQKFRANRSMKQARQBFVLRTIATV 1541
Db 1477 -----RKFRBEEQLRQERBEEQLRQERDRKFRBEEQLLQERBEEQLRQ 1507
QY 1542 HLQKFRGKRLMIE 1555
Db 1508 -VWEEDKRRQVIE 1520

```

```

RESULT 3
MY5A CHICK
ID MY5A CHICK STANDARD; PRT; 1829 AA.
AC Q02440.
DT 01-JUN-1994 (Rel. 29, Created)
DE 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
heavy chain p190) (Myosin-V).
GN MYO5A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93012002; PubMed=1383040;
RA Sanders G., Lichte B., Meyer H.E., Kilian M.W.;
RT "cDNA encoding the chicken ortholog of the mouse dilute gene product.

```

```

RT RT
RL FEBS Lett. 311:295-298(1992).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC TISSUE=Brain;
RX MEDLINE=93107155; PubMed=1469047;
RA Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
de Camilli P.V., Larson R.E., Mooseker M.S.;
RT "Primary structure and cellular localization of chicken brain
myosin-V (p190), an unconventional myosin with calmodulin light
chains."
RL J. Cell Biol. 119:1541-1557(1992).
CC -1- FUNCTION: Processive actin-based motor that can move in large
steps approximating the 36-nm pseudo-repeat of the actin filament.
May be involved in melanosome transport, or alternatively, it may
be required for some polarization process involved in dendrite
formation (By similarity).
CC -1- SUBUNIT: May be a homodimer, which associates with multiple
calmodulin or myosin light chains.
CC -1- SUBCELLULAR LOCATION: Golgi-derived cytoplasmic membranes
(potential).
CC -1- TISSUE SPECIFICITY: Neuronal and nonneuronal cells of the brain.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 dilute domain.
CC -1- SIMILARITY: Contains 6 IQ domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67251; CAA47673.1; -.
DR EMBL; Z11718; CAA77782.1; -.
DR PIR; S19188; S19188.
DR HSP; P10587; 1BR2.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSTINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 6.
DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
Coiled coil; Phosphorylation.
KW Coiled coil; MYOSIN HEAD-LIKE.
FT DOMAIN 1 766
FT DOMAIN 767 789
FT DOMAIN 790 814
FT DOMAIN 815 837
FT DOMAIN 838 862
FT DOMAIN 863 887
FT DOMAIN 888 915
FT DOMAIN 915 1239
FT DOMAIN 1239 1419
FT DOMAIN 1419 1661
FT DOMAIN 1661 1766
FT NP BIND 163 170
FT DOMAIN 644 666
FT MOD RES 1734 1734
FT CONFLICT 1142 1142
FT CONFLICT 1829 1829
SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

```

Query Match 3.3%; Score 317.5; DB 1; Length 1829;  
Best local similarity 18.3%; Pred. No. 8.9e-10;  
Matches 351; Conservative 271; Mismatches 670; Indels 623; Gaps 82;





Db 621 RPPASGDAQEQKQVHPGSHVSELTOTETEEDSS-----GETLSQRLERLEL 671  
QY 773 VLQHQRTQDEFD-YAFNNLAVDLR-----DGVRLTRVVEVILLRDDLTRQLRPAISR 825  
Db 672 FTQGERGEVLKLEQSAENLQARLEBSSSLQLQREKREKDLISSTSTSNLGOELSS 731  
QY 826 LQRI FNVKLALGALGANFQAGDIAAQDIVDGHREKTLSLWLIQYKF-----RSPKFH 880  
Db 732 MQRSESVATLDA-GE-----GGPVLFEKC-----EKSLSKLNSELSEAYRKANDROAKEN 780  
QY 881 AAATVLQKWRH-----WLHVVIQRIPIHKELMRHRAATVIQAVF-----RGHQMKY 930  
Db 781 VSKLAKAKNCHTQSLSELHKV-----KEASTAVETVIVVETVAVTAPNGKALABY 833  
QY 931 VKLFXTERTQAAIILQKFTERRYLAKOLYQSYHSIITI-----QRWRAQQLGRQHRQP 985  
Db 834 BOL-NAQNAELKANVIRLROELDELESPEPTEAPLAIVGSOSQREDEILQOSQ-----887  
QY 986 VELREAAIFLQIRWRRILFAK-----KLAAAEATARIQSQKQAAASYIQMWRTY-QLG 1040  
Db 888 ---LEDARSLQAEQOQIBEQVDQIKELRQTEAQQLVARQSAITQLQSEQFDOLL 944  
QY 1041 RIQREFTLRQDLINFEVORMKSKSMLEQR-----KEFOQL-----KRAAINIQ 1085  
Db 945 NSKEMSEKOLEQOQTRIRRELEAREASLEGELSILQTLVAEQKQQLIESVSEHALNLK 1004  
QY 1086 --QRWRKSLMRKCA-----DVL--ALR--SSVLKVQAVRKATIOWRIDENHYYSRKVNI 1136  
Db 1005 MLELQSAQELRELRAKEDPDOLRALRYSKSLVAQOQVRELT-----SSQETVD 1053  
QY 1137 CLOQLRALI-----MKMREQENYLRNLNASILVOKRYMR-----QQMIDRNAY 1182  
Db 1054 ALNQIOEQYQGLEHAHKEEQFNELRB---KLKVALNLKKTQDNADLEQKVDELISQ 1110  
QY 1183 LRTRKCI-----NVQRWRATLQMRERKNYHLHQTTKYRIQIKFRKR 1227  
Db 1111 LQEQQLVKQKEEVEREPIDVNRHVEQLQOQVSKLNEDLKAKIHLNLR-----DALR 1164  
QY 1228 EMKQORAFLOLKKVTLVQKRRAL-----LQMKER-----QEVYHLREV 1269  
Db 1165 QLKQIQIQEQQL-----IQERDAELOANILVSELREERQAEQVFPOLQGENSLREE 1218  
QY 1270 TIKLQRRPH-----AQKSMFMEAK--YRGTQAAVSLQW-----1302  
Db 1219 ISKLEETHNLQGVNEBPTAVEDLRQLEAKSKFKFKELIKLRNATIQSLQELQOL 1278  
QY 1303 -----HWRN-----HLIRKERNFLOLRQAAITLQRYRA-----RLNMIKOLKS 1343  
Db 1279 QOQDQSEVHVNRANAHEQLRLKEDAEITALRQELIKLERSRAAGEGDDTITKTHQLL 1338  
QY 1344 YAQLKQAAITITQYRAKAMQKVLYQKQREALIKVORYVRGNLEMRKQIEVYQKORQ 1403  
Db 1339 ESQSQQASQVARELQQLRVOLTAQEQHALLAQOYASDKANFEM-----1386  
QY 1404 AVIRLQKWRMSIRDMWLCKAGY-----RRRLSSLSIQKWRATVQARRQREIFLSTIR 1457  
Db 1387 TIARLETHGEOAKIQEDASYIESLEAQNTLEQARSALAEQASQANQO-----AASQD 1442  
QY 1458 KVRMLQAFIRATLQMRQREFFEMKRRAAVVIQRRF-----1493  
Db 1443 KVQILQEQ-----QKQEQEQEQKQOQOQQLQERFVLEQREQAQSQLELLTSEABES 1496  
QY 1494 RARCAMLKARQDYOLQSSVILVOKKFRANSMKQARQEFVQLR-----TIAVHLQKQF 1547  
Db 1497 RQJAGLRTEYSLAKHSQLTATAQAREQMSHQSQELAEALRQQLDVKEADLHRQOV 1556  
QY 1548 RQKRLMIEQRCNFOLLRCMFGQARARG-----FMARKRFOALMTPEMDLIRQKAAK 1602  
Db 1557 YDAKLAQATFELDE-LECDLNSHVERRAAETREICQQLERSQELVAQRTBELQRLNEEFQ 1615  
QY 1603 VIQRYWRYGLIRR-----RQKH-----QGLIDIRKRIQALRQAKAVNSVRCKV-----Q 1647

Db 1616 EVERE-RSTLSREVTLIRLQHDSEAQDVLEQLRQMAQDKTEMNLRTOIDALCANHS 1674  
QY 1648 EAVRFLRGFIASDALAVLSQDLRLSRVTPHLLMCSEFPMSTFCYGIMAAQAIRSEVDKQL 1707  
Db 1675 OSLOALQORIAELDTIGQNTDDQV-----YIETE-NKRL 1708  
QY 1708 IERCSTRIILNLAR 1720  
Db 1709 AEQLSELQOQLAR 1721  
RESULT 5  
MY5A HUMAN  
ID MY5A HUMAN STANDARD; PRT; 1855 AA.  
AC Q9Y4I1; O60653; Q07902; Q16249; Q9UE31; Q9UE31;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)  
DE (Myosin heavy chain 12) (Myosin).  
GN MYOSA OR MYH12.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Skin;  
RA Meurers B.H., Zimmermann R., Vosberg H.P.;  
RT "The complete cDNA for human myosin heavy chain 12, a class V  
myosin."  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.  
RX MEDLINE=97351514; PubMed=9207796;  
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,  
Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;  
RT "Griscelli disease maps to chromosome 15q21 and is associated with  
mutations in the myosin-Va gene."  
RL Nat. Genet. 16:289-292 (1997).  
RN [3]  
RP ERRATUM.  
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,  
Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;  
RL Nat. Genet. 23:373-373 (1999).  
RN [4]  
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).  
RC TISSUE=Fetal brain;  
RX MEDLINE=94245227; PubMed=8188282;  
RA Engle L.J., Kennett R.H.;  
RT "Cloning, analysis, and chromosomal localization of myosin (MYH12),  
the human homologue to the mouse dilute gene."  
RL Genomics 19:407-416 (1994).  
RN [5]  
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).  
RC TISSUE=Brain;  
RX MEDLINE=95136715; PubMed=7835087;  
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,  
Jenkins N.A.;  
RT "Cloning and regional assignment of the human myosin heavy chain 12  
(MYH12) gene to chromosome band 15q21."  
RL Cytogenet. Cell Genet. 69:53-58 (1995).  
RN [6]  
RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).  
RA Edgar A.J., Bennett J.P.;  
RT "Inhibition of dendrite formation in melanocytes transiently  
transfected with antisense DNA to myosin V."  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP FUNCTION.  
RX MEDLINE=99376094; PubMed=10448864;  
RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,  
Cheney R.E.;



```

Db      886 M---HAIYLOCCFRMMAKRELUKUKIAR-----SVERY 918
QY      1046 EPLR--ORDLIMFVQRRMSKWSMLQKREFOOLKRAAINIQORWRAKLSMRKCNADYLA 1103
Db      919 KKLHIGWENIMQKQVD-----BQNDKYCLVEKLNLEGYINSETEKLSDLERLQ 972
QY      1104 LRSVLKQVAYRATQMRID--RNHYISLRKQVCLQORLRAIMQRRQRENYLRLNA 1161
Db      973 LSEEEAVATGRVLSQEEIAKRLKOLEQTRSEKKCIEEHADRYKQETEQVLSNLKEEN- 1031
QY      1162 SILVQRYRMEQMIODENAYLTRKCIINVRWRATLQMRERKNYLHLQTTTKRIOI 1221
Db      1032 TLXQKEALNHRIVQAKENTMEKKL-VETKQLELDLNDERLYQNLLNEFSRLEE 1090
QY      1222 KFR-----AKREMKQRAAFQLQKVT- 1243
Db      1091 RYDDLKEEMTLMVHPKPGHKRTDSTHSSNESEYIFSSIEAEMEDIPSRTEEPSEKKVPL 1150
QY      1244 ----LVVQKERRALLQMRKERQBYLHLREVITIKLQRRFHAQKSMFRMAKYRGTOAAYS 1298
Db      1151 DMSIFLKIQKRVTELEQEKQVMQDELDRKEEQVL-----RSKAKEERPPQIRGALEVE 1204
QY      1299 CLOQWNRNHL--LRKBERNSFLQRAAITLQRR-----YRARLANMIKQLKSY 1344
Db      1205 SLK---RQELSESENKKLKNELNLRKA---LSEKSAPEVTPGAPAVRVLMELQTSVSEE 1258
QY      1345 AQLKQAAITQTRYAKAMQKVVL-----YQKOR-----BAIKV 1381
Db      1259 LDVRKEEVLII--LRSQVSKQEAIPKDKNTMTDSTILLEVDQVKMKQGETAQAYIGI 1315
QY      1382 ORRVGNL-----EMRKQIE--VYQKORQA---VIRLOKWRISRD-----MRLC 1421
Db      1316 KETRNSALDYHNEDECELWLVEGLKQANRLLESQLOSKSHENAEALRGETQSLK 1375
QY      1422 KAGYRIRLSLSIQKRWATVQARRQREIFLSTIRKVRMQAFIRATILMRQORREFEM 1481
Db      1376 EENNROQQLLAQNQLPPEARIEASLQHEITRLTNENMLDMEQ-----LEKQDKTVRKL 1429
QY      1482 KRRAAVIQRRFRARCAMLKARQDYQLIQSSV--ILVQKFRANRSMKQARQEFVOLRTI 1539
Db      1430 KQQLVFAKKIGILEVQGMENISFGQIIDEPIRPVNIPEKEDFOGMLKYKKEDEOKLVK 1489
QY      1540 AVHLQQRFRGKRLMIQORNCQFOLLRCSPMGFORARANGFAR-----KRFQALMTPEM 1591
Db      1490 NLILKLPKRGAVNL-----IPGLPAVILPWCVRHADYLDNDQKVESLITSTI 1537
QY      1592 MDLIRQRAAKVIR-----YRQGYLIR-----RRQKHQGLL 1623
Db      1538 NSI-----KKVLKRGDDFETVSPWLSNTCRFLHCLKQYSGEGFMKHNTRSQRNEHCLT 1591
QY      1624 DIRKRIQLRQKAKVNSVRCKV-QEAVRFLRG---RPIASDAL-----1663
Db      1592 NF--DLAEYIQ---VLSDLAIQIYQQLVRLVLENILQPMIVSGMLEHETIQGVSQVPTGL 1646
QY      1664 --AVLSQDLRLSRTVPHLLMWCSEFMSTFCYGINAQARISEVDKQLIERCSRILN----1717
Db      1647 RKTSTSIADGETVTLDSILRQLNSFHSVMQGHGMDPELIRQVVKQPMFYIITGATLANLL 1706
QY      1718 -----LARVNSTTV-----NTQEGG-----LVTTAQMIL--LRWCDDKDEI 1751
Db      1707 RDMCWSKSGNQIRYNVSQLEEWLRDNLMNSGAKETLEPLIQAAQLLQVKKKTDDDAEA 1766
QY      1752 FNTLCTILI 1759
Db      1767 ICSMCNAL 1774

```

RESULT 6

ID\_PLE1 RAT

AC P30427; 008879; 008880; 008881; PRT: 4687 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Plectin 1 (PLTN) (PCN)
GN      PLECI.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM 1).
RX      TISSUE=Glial tumor;
RX      MEDLINE=91268156; PubMed=2050743;
RA      Wiche G., Becker B., Lubet K., Weitzer G., Castanon M.J.,
RA      Hauptmann R., Stratowa C., Stewart M.;
RT      "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT      chain with a three-domain structure based on a central alpha-helical
RT      coiled coil.";
RL      J. Cell Biol. 114:83-99 (1991).
RN      [2]
RP      REVISIONS.
RX      TISSUE=Glial tumor;
RX      MEDLINE=96210632; PubMed=8633055;
RA      Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT      "Human plectin: organization of the gene, sequence analysis, and
RT      chromosome localization (8q24).";
RL      Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283 (1996).
RN      [3]
RP      PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4), AND TISSUE
RP      SPECIFICITY.
RX      TISSUE=Glial tumor;
RX      MEDLINE=97321050; PubMed=9177781;
RA      Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA      Wiche G.;
RT      "Plectin transcript diversity: identification and tissue distribution
RT      of variants with distinct first coding exons and rodless isoforms.";
RL      Genomics 42:115-125 (1997).
CC      -!- FUNCTION: Interlinks intermediate filaments with microtubules and
CC      microfilaments and anchors intermediate filaments to desmosomes or
CC      hemidesmosomes. May be involved not only in the crosslinking and
CC      stabilization of cytoskeletal intermediate filaments network, but
CC      also in the regulation of their dynamics.
CC      -!- SUBUNIT: Homodimer or homotetramer.
CC      -!- ALTERNATIVE PRODUCTS;
CC      Event=Alternative splicing; Named isoforms=4;
CC      Name=1;
CC      IsoId=P30427-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=P30427-2; Sequence=VSP_005050;
CC      Name=3;
CC      IsoId=P30427-3; Sequence=VSP_005051;
CC      Name=4;
CC      IsoId=P30427-4; Sequence=VSP_005052;
CC      -!- TISSUE SPECIFICITY: Widely expressed with highest expression in
CC      skeletal muscle and lowest in thymus.
CC      -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with
CC      vimentin, desmin, GAP, cytokeratins, lamin B; whereas both the N-
CC      and the C-terminus can bind integrin beta-4.
CC      -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC      INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC      -!- SIMILARITY: Contains 1 actin-binding domain.
CC      -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC      -!- SIMILARITY: Contains 33 plectrin repeats.
CC      -!- SIMILARITY: Contains 4 spectrin repeats.
CC      -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC      -!- CAUTION: Isoform 4 is a fragment at the N-terminus.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----

```







```
Db 1703 VOVALETAQSAEVELQSKRSPFAKTAQALERTLOBEHVTVTQLREBAERRAQQQAEEAER 1762
QY 697 L7EYVAETIKCHSL---OKILFLPLPFDQAKOKRIIVKHNPCFLPVKSPKSTKDWIILLRPS 753
Db 1763 ARREARELERWOLKANEALRLRLQAEVVAQKSLAQ-----ADAEKQKE-----1807
QY 754 SELLANIGDITRELRRLGVVLQH--RQTFIDEPDYAFNNLAVDLRDGVRLTRVVEVILLR 811
Db 1808 -----BAEREAARRGAEQAVQRELAEQELEKQQLTEGTAQQRLLAAEQELILRLR 1859
QY 812 DDL-----TROLRVPALSLQ-----RIFNVKLAIGALGANFOLGGDIAAQDITV 856
Db 1860 AETEQEHOROLLEELABLOHEATAATOKQOELEABLA-KVRAEMEVILLASKARAEBS 1918
QY 857 DGRHKRTLSLLWLIYKFRSPKPHAAATVLOKWRHRLHVLIORIRKELMRHRAAT 916
Db 1919 RSTSEKQRLBAEAGFRELAEAR-----LRAEAERARRHRELAEDDAARQ 1967
QY 917 VIQAVFRGHOMKVKYKLFKTER--TOAIILOKFTF-----RYLAQKOLYQSYHSIITIQ 969
Db 1968 RAEA--DGVLTEKLAASISATRLKTEABIALKEKAENERLRLAEDFAQRR-----2018
QY 970 RWRRAQQLGROH-----QRFVELREA-----IFLQIRWRRRLFPAKLLAAAEYAR 1016
Db 2019 ---RUEQAAQKHADIEERLQALRKASESELERQKGLVEDTLRORRQVEEIMAL-----2070
QY 1017 LQSKQQAASAYIQWRTYQGLRIQ-----RHEFLRQDLIM-----1055
Db 2071 --KASPEKAAAGKALE---LELGRIRSAEDTMRKSELAEQAEARQRLAAEHEQRRE 2125
QY 1056 ---FVORMRKSWMLBQK-----EFQOLKRAAINIQFWRAKLSMRKKNADYLAIRSSV 1108
Db 2126 AEERVQSLAAEEAARQKVALEVERLK-AKVEARELRERAEQESARQOLQAEAAQ 2184
QY 1109 LKQVAYRATQMRIDRHHYSLRKNVLCLOPLRAIMWRQRENYL-RUN-----1160
Db 2185 KRLQAEKA-----HAFVVOQREBELQOTLOEQBNMLERLASEAAAR 2228
QY 1161 -ASIIIVKRYMRQOMIQDRNAYLTRKCIINVORRWATLO-----MRERKNYLHL 1212
Db 2229 AAEBAEAREQAEARASQKQVEAEARLKQSAEQQAQQAQAAAKLRKEAB-----2283
QY 1213 QTTKRIQIKFRAKREMKQABFLOLKVVTLVQKRRALLQMKERQEVYHLHREVTIK 1272
Db 2284 QEAARRAQEAQALQKQAADAEMKHKKFA-----EOTLRQKAQVQELTTLR--LQ 2334
QY 1273 LORRPHACKSM---RPMKAYRGTOAAVSCLO-----MHWNRHL 1308
Db 2335 LEETHD-QKSILDBELQRLKAEVTEAARQSQVEBELFSVRVQMBELGKLKARIBAEENRA 2393
QY 1309 LKREERN---FLQRLQAAITLQRRYRARLNMKOLKSYAOLKQAA---ITIQTRYRAKKA 1363
Db 2394 LILRDKONTQRFLEBAEKMKQVABEAARLSVAQ--EAARLQLAEEDLQORALAEKM 2451
QY 1364 MQQVVLVQ-----KQRAIIKVORRYKGNLEMQRQIEVYQKQAVIRLOKWRSI---1415
Db 2452 LKEMQAVQEAATRLKAEAELEQQOKELAQEABRLQADKEQMAQQLVEETQGFORTLEAE 2511
QY 1416 --RDMRLCKAGYR-RIRLSLSISIQKRWATVQARRQRE-----IFLSTIRKURL 1461
Db 2512 RQRLQESAEABRLKRLNAENS--RAQAEABEDAQRFKQABEEIGKHLRTELATOQKVTL 2570
QY 1462 MQAFIRATL-LMRQOR-REFEMKERRAAVVIOK-----RFRARCAMLKARQDYQLIOSSVI 1514
Db 2571 VQ-----TLETQOSQDQAEARLEAIAELEREKULKQBAKULQKLS-EMQTVQOQBI 2624
QY 1515 L-----VQKFRANKSMQAPQEFVQLFTIAVHLQKFE-----RGRKRLMEQRNCFOLLR 1564
Db 2625 LQETQALOKSPLEKSDLSLQREFTIQE--NAKLEQLFQDEVAKAKOLQEEQ-----2674
QY 1565 CSMPFGQARAGFWARKRFQALMTPFEMDLI-----RQKRAKVITQVRVGY--LIR 1614
Db 2675 -----QRQQQNEQEQKQELVASMEEAARRRQREAEQGVRRKQEBLQRLQE 2718
```

```
QY 1615 BRQKHGCHL-----DIRKRIALRQPAKA 1638
Db 2719 QRQQQEKLLAEENQRRLERLQRLLEEHR 2747
```

## RESULT 7

```
MYSA RAT
ID MYSA RAT STANDARD; PRT; 1828 AA.
AC Q9QYF3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
GN MYOSA.
OS Rattus norvegicus (Rat).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RA MEDLINE=20380823; PubMed=10920234;
RA Futaki S., Takagishi Y., Hayashi Y., Ohmori S., Kanou Y., Inouye M.,
RA Oda S., Seo H., Iwakawa Y., Murata Y.;
RT "Identification of a novel myosin-Va mutation in an ataxic mutant
RT rat, dilute-opisthotonus.";
RL Mamm. Genome 11:649-655(2000).
CC -!- FUNCTION: Processive actin-based motor that can move in large
CC steps approximating the 36-nm pseudo-repeat of the actin filament.
CC Involved in melanosome transport. May also be required for some
CC polarization process involved in dendrite formation (By
CC similarity).
CC -!- SUBUNIT: May be a homodimer, which associates with multiple
CC calmodulin or myosin light chains. Binds MLPH and MYRIP (By
CC similarity).
CC -!- DISEASE: Defects in MYOSA are a cause of Dilute-opisthotonus
CC (dop). Dop rats have diluted coat color and are occasionally
CC associated with severe neurological disorders.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 6 IQ domains.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB035736; BAA88350.1; -.
CC HSSP; P10587; 1BR2.
CC InterPro; IPR002710; DIL.
CC InterPro; IPR000048; IQ region.
CC Pfam; PFO1843; DIL; 1.
CC Pfam; PFO0612; IQ; 6.
CC Pfam; PFO0063; myosin head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD003376; DIL; 1.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 6.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS00096; IQ; 6.
CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Phosphorylation.
CC DOMAIN 1 765 MYOSIN HEAD-LIKE.
CC DOMAIN 766 788 IQ 1.
CC DOMAIN 789 813 IQ 2.
CC DOMAIN 814 836 IQ 3.
CC DOMAIN 837 861 IQ 4.
CC DOMAIN 862 884 IQ 5.
CC DOMAIN 885 914 IQ 6.
```

FT DOMAIN 914 1239 COILED COIL (POTENTIAL).  
 FT DOMAIN 1314 1418 COILED COIL (POTENTIAL).  
 FT DOMAIN 1660 1765 DILUTE.  
 FT NP BIND 163 170 ATP (POTENTIAL).  
 FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).  
 FT MOD RES 1733 1733 PHOSPHORYLATION (POTENTIAL).  
 SQ SEQUENCE 1828 AA; 211762 MW; 583DEIC89AE36123 CRC64;

Query Match  
 Best Local Similarity 3.1%; Score 295.5; DB 1; Length 1828;  
 Best Local Similarity 19.4%; Pred. No. 1.5e-08;  
 Matches 287; Conservative 231; Mismatches 538; Indels 427; Gaps 68;

QY 559 KLENEV--RNKELVAPTEBQSMVYLTKYRLTLKAAVELFSPQMRPLPCSKVAVYN 616  
 DB 410 KLFNVRVGNVQALHSAVKHSTGIVLDIYGFTEFINSPEQF-----CIN---YAN 458  
 QY 617 KQALRIRSDNLHL-----DVVMQRTILELL-----CFNPLMLRLGLVGVFGKIOM 664  
 DB 459 E---KLQQFNHVFVLEQEEYKKEQIPWTLIDFYDNQPCINLIESKGLDLDLDECKM 515  
 QY 665 QSNRDIVGLSTFNLRLFRNKC---BEQVYS-KAY-----TLT 698  
 DB 516 PKGTDDTWAQKLYNTHL--NKCALFEXPRMSKAFIKHFPADKVEYQCEGFLEKNKDVF 573  
 QY 699 EYVARTIKKHSLOKILPLFLDOAKOKRIVK-----HNPCLFVKKSP----- 741  
 DB 574 EQIKVLKSKFK-----MLPELFQDDEKALSPSATSSTGTPITRPVVKTKGPGQAK 629  
 QY 742 -HKETKDILLFSELLANIGDITRELRRLGYVLQHRQTFDLDFDFAFNLANVLDLRGVR 800  
 DB 630 EHKKTGVLQFNSLHLL-----METLNATPHVYCIKPNDEKF-PFTFDEKRAVQ 679  
 QY 801 LTRVVEVILLDDLTROLRVPA-----ISLQRI FNVKLALG----- 837  
 DB 680 QLRACGYL-----ETIRISAGFSPGRWTYQEFSSRYVLMKQKQVLDGRKOTCONVLE 732  
 QY 838 --ALGEANFOLGGDIAAQDIVDGHREKTLISLLWLIY--KFRSPKFAAAATVLOK---W 889  
 DB 733 KLILDKKYQFG-----KTKLFFRAGQVAVLEKLADKLRAACIRIQKTIRGW 780  
 QY 890 -WRRHWL-----HVVIQRRIR-----HKELMRHRRAATVIAQVFRGHQMRKVKVLEKTE 937  
 DB 781 LLRKRYLCMORAATVQRYRGYQARCYAKFLRRTKAATTKYRWYVVRKYKI----- 836  
 QY 938 RTQAAIILQKTRYLAKQILQSY--HSLITIQ---RWRAQOLGRQHRQVRLREA 991  
 DB 837 RRAATVILQSYLGYLARNR-YRKILREHKAIVIOKVRGW-----LARTHYKRTM---KA 888  
 QY 992 AIFLQIRWRRLFAKLLAAAEATARLQSQKQAAASYIQMWRTYQLGRITQRHEFLR-- 1049  
 DB 889 ILYIQCFPRMMAKRELKLEAR-----SVERYKKLHIG 924  
 QY 1050 QRDILMFQRRMRSKWMLQKRFQOLKRAINIQQWRRAKLSMRKNADYIALRSSVL 1109  
 DB 925 MENKIMQQRKVD-----EQNDKYKCLMEKLTNLEGVYNSETEKLRNDVRELQLSBEEA 978  
 QY 1110 KVQAYRATIQ-----WRIDENHYLSRKVNICLQRLRAIMKWRQRENYLRLNASIL 1164  
 DB 979 KVATGRVLSQEEIAKRLKDLQETREKKST---EERADKYQTEQLVSNLKEEN-TLL 1034  
 QY 1165 VQKRYRQRMIOQRNAVLR--RKCIINVQRWRATIQMERERKNVLIHLOTTTKRIQIK 1222  
 DB 1035 KQKETLNHLNVAQAKETETWERKL-----VETKQLELDLNDERLYQNILNFSRLBER 1091  
 QY 1223 F-RAKREMKQ-----RAEFLQI-----KKVT-- 1243  
 DB 1092 YDDLKEEMTLNVPKPGHKRTDSTHSSNESEYTFSEFAETDIAPTEBTEKKVPLD 1151  
 QY 1244 ----LVVQKRRALLOMERKEQVYLHREVITKLQRRFHAOKSMFRMAKVGTAQAVSC 1299  
 DB 1152 MSFLKQKRVTELGQEQKLMODELDRKEEQVL-----RSKAKGGERPQIRGABLGYES 1205  
 QY 1300 LQMEWRNHL--LRKERNSTFLQROAAITLQRR-----YRRLANMIKOLKSYA 1345

DB 1206 LK---RQELSENKKLKNELNLRKA---LSKSAPEVNAPGAPAYRVLMSQUTAVSEEL 1259  
 QY 1346 QLKQAAITITQTRYRAKAMQKQVLYQKOREALI-----KVORRYRGNL 1389  
 DB 1260 DVRKEVLI---LRSQVSKQKAIQPDKNKTWTSTILLEDVQKMKDGETAQAVIGLK 1316  
 QY 1390 EMRKOIEVYQKQOAVIRLQKWRMSIRD-----MELCKAGYRRIRLSSLSIQKWR 1440  
 DB 1317 ETNRLLE-----SQLSQKRSHEAEALRGEISLKENNQQQLLAQNLQPPPE 1367  
 QY 1441 ATVQARRQREIFLSTIRKVRILMQAFTRAVILMRQORREFEMKERRAAVITQRRFRACML 1500  
 DB 1368 ARTEASLQHEITRLTNTNENLDLMEQ-----LEKQDKTVRKLLKQKLVFAKKIGELVVGOM 1421  
 QY 1501 KARQDQLIQQSV--TLVORKFRANSMKQARQEFVQLRTIAVHLOQKFRGKEL-MIEQR 1557  
 DB 1422 ENISPGQIIDEPIRPVNIIPRKGDFOGMLYKREDEQKLVKNLILLEKPRGVAVNLISGL 1481  
 QY 1558 NCFQLLRCSMPGQFQARGFQ-ARKRFOALMTPMEMDLIRQKRAAKVIOQ-----Y 1607  
 DB 1482 PAVILFWC-----VRHADYLDQDKVRSLLTSTINSI-----KKVLKRGDDFTVSVF 1529  
 QY 1608 WRGYLIR-----RRQHQGLDIRKRIARQLOEAKAVNSVRCKV-Q 1647  
 DB 1530 WLSNTRCFHLHCKQYSGEFGMKHNTSRQNEHCLTNF--DLAEYRQ---VLSDLAIQIYQ 1584  
 QY 1648 EAVRELRG--RFIASDAL-----AVLSQDLRLSRTVPHILMWCSEF 1686  
 DB 1585 QLVRVLENILQPMIVSGMLEHEFIVQGVGVKPTGLKRTSSIADEGTYTLDLSILQLNSF 1644  
 QY 1687 MSTFCYGINAQAIRSEVDKQILERCRIILN-----LARNSTTV----- 1726  
 DB 1645 HSNVCQHGMPELIKQVVKQMFYIVGAILTNLLLRKDMGSKMQIRYNVSQLEEWLR 1704  
 QY 1727 --NTFOEGG-----LVTIAQML--LRWCKDKSEIFNLCTLI 1759  
 DB 1705 DKNLMNSGAKETLEPLTQAAQLQVKKKTDDBAICMCNAL 1747

RESULT 8  
 PLE1\_HUMAN STANDARD; PRT; 4684 AA.  
 AC Q15149; Q15148; Q16640;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Plectin 1 (PLTN) (FCN) (Hemidesmosomal protein 1) (HD1).  
 GN PLECI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 RT chromosome localization (8q24).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283 (1996).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
 RX MEDLINE=96312447; PubMed=8698233;  
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,  
 RA Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,  
 RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,  
 RA Uitto J.;  
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:  
 RT cDNA cloning and genomic organization.";  
 RL Genes Dev. 10:1724-1735 (1996).  
 [3]  
 RN VARIANT MD-BES 1003-GLN--ALA-1005 DEL.

MEDLINE=97049959; PubMed=8894687;  
 RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
 RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.,  
 RA "Homozygous deletion mutations in the plectin gene (PLEC1) in patients  
 RT with epidermolysis bullosa simplex associated with late-onset  
 RT muscular dystrophy.";  
 RL Hum. Mol. Genet. 5:1539-1546(1996).  
 [4]  
 RP VARIANT MD-EBS LEU-429 INS.  
 RX MEDLINE=21090821; PubMed=11159198;  
 RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,  
 RA Mass W., Hametner R., Klaussegger A., Huber A., Pohla-Gubo G.,  
 RA Wiche G., Uitto J., Hintner H.,  
 RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
 RT the plectin gene causes epidermolysis bullosa simplex with plectin  
 RT deficiency.";  
 RL Am. J. Pathol. 158:617-625(2001).  
 [5]  
 RP VARIANT EBS1 TRP-2110.  
 RX MEDLINE=21841370; PubMed=11851880;  
 RA Koss-Harnes D., Hoeyheim B., Anton-Lamprecht I., Gjesti A.,  
 RA Joergensen R.S., Jahnson F.L., Olaisen B., Wiche G.,  
 RA Gedde-Dahl T. Jr.,  
 RT "A site-specific plectin mutation causes dominant epidermolysis  
 RT bullosa simplex Ogna: two identical de novo mutations.";  
 RL J. Invest. Dermatol. 118:87-93(2002).  
 CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and  
 CC microfilaments and anchors intermediate filaments to desmosomes or  
 CC hemidesmosomes. Could also bind muscle proteins such as actin to  
 CC membrane complexes in muscle. May be involved not only in the  
 CC crosslinking and stabilization of cytoskeletal intermediate  
 CC filaments network, but also in the regulation of their dynamics.  
 CC -!- SUBUNIT: Homodimer or homotetramer.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q15149-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q15149-2; Sequence=VSP\_005030;  
 CC Name=3;  
 CC IsoId=Q15149-3; Sequence=VSP\_005030, VSP\_005031;  
 CC -!- TISSUE SPECIFICITY: Widely expressed with highest levels in  
 CC muscle, heart, placenta and spinal cord.  
 CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
 CC vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N-  
 CC and the C-terminus can bind integrin beta-4.  
 CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa  
 CC simplex with muscular dystrophy (MD-EBS) [MIM:226670]; an  
 CC autosomal recessive disorder characterized by epidermal blister  
 CC formation at the level of the hemidesmosome and associated with  
 CC late-onset muscular dystrophy.  
 CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa  
 CC simplex 1 (EBS1) [MIM:131950]; also called epidermolysis bullosa  
 CC simplex Ogna type. EBS1 is an autosomal dominant form of  
 CC epidermolysis bullosa simplex differentiated from the more  
 CC generalized form of Koeber [MIM:131900] and the localized form of  
 CC Weber and Cockayne [MIM:131800] by the occurrence of skin  
 CC bruising.  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 33 plectrin repeats.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -!- SIMILARITY: Belongs to the plectin or cytolinker family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; Z54367; CA91196.1; -  
 DR EMBL; U53204; AAB05427.1; -  
 DR EMBL; U63610; AAB05428.1; -  
 DR EMBL; U63609; AAB05428.1; JOINED.  
 DR EMBL; X97053; CA65765.1; -  
 DR FIR; CS9404; A59404.  
 DR HSSP; Q01082; LBKR.  
 DR Genew; HGNC:9069; PLEC1.  
 DR GK; Q15149; -  
 DR MIM; 601282; -  
 DR MIM; 236670; -  
 DR MIM; 131950; -  
 DR GO; GO:0008307; F:Structural constituent of muscle; TAS.  
 DR InterPro; IPR001589; Actbind\_actinin.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR001101; Plectin repeat.  
 DR InterPro; IPR005326; S10 plectin\_N.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00881; Plectin; 19.  
 DR Pfam; PF03501; S10 plectin; 1.  
 DR ProDom; PD006662; S10 plectin\_N; 1.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00250; PLEC; 34.  
 DR PROSITE; PS00019; ACTININ\_1; FALSE\_NEG.  
 DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
 KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
 KW Disease mutation.  
 FT DOMAIN 1 1470  
 FT DOMAIN 1471 2755  
 FT DOMAIN 2756 4684  
 FT DOMAIN 175 400  
 FT DOMAIN 179 282  
 FT DOMAIN 295 397  
 FT REPEAT 645 710  
 FT REPEAT 740 824  
 FT REPEAT 837 930  
 FT REPEAT 1315 1415  
 FT DOMAIN 1469 2756  
 FT REPEAT 2826 2863  
 FT REPEAT 2864 2901  
 FT REPEAT 2902 2939  
 FT REPEAT 2940 2977  
 FT REPEAT 2981 3015  
 FT REPEAT 3116 3153  
 FT REPEAT 3154 3191  
 FT REPEAT 3192 3229  
 FT REPEAT 3230 3267  
 FT REPEAT 3268 3305  
 FT REPEAT 3306 3343  
 FT REPEAT 3485 3522  
 FT REPEAT 3523 3560  
 FT REPEAT 3561 3598  
 FT REPEAT 3599 3636  
 FT REPEAT 3640 3674  
 FT REPEAT 3820 3857  
 FT REPEAT 3858 3895  
 FT REPEAT 3896 3933  
 FT REPEAT 3934 3971  
 FT REPEAT 3975 4008  
 FT REPEAT 4061 4100  
 FT REPEAT 4101 4138  
 FT REPEAT 4139 4176  
 FT REPEAT 4177 4214  
 FT REPEAT 4218 4252  
 FT REPEAT 4265 4305  
 FT REPEAT 4319 4356  
 FT REPEAT 4408 4445  
 FT REPEAT 4446 4483  
 FT REPEAT 4484 4521

```

FT REPEAT 4522 4559 PLECTIN 32.
FT REPEAT 4560 4537 PLECTIN 33.
FT DOMAIN 4250 4300 BINDING TO INTERMEDIATE FILAMENTS
      (BY SIMILARITY).
FT DOMAIN 4625 4640 4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT MOD_RES 4539 4539 PHOSPHORYLATION (BY CDC2)
      (BY SIMILARITY).
FT VARSPLIC 1 174 NYAGMLMPDQRLALRYEVLFRGVAVAKDRPRSLHHPV
      GVTNIGVMAASLRARGLVRETFAWCHFFFWLTNEGIAHL
      RYLLHLPPELVAASLQRRVRVAMVPARTPHVQVQGL
      GSPPKRGLPTLEQLYRKELEBEVSPETPVVPAQTTLA
      RPPGPAPAT -> MSGEDAENAVSDEVSNSSGSPSPGD
      TLPNLGKTQGRSGGAGSGSVLDPARAVIRIA (in
      isoform 2 and isoform 3).
      /FTid=VSP_005030.
      Missing (in isoform 3).
      /FTid=VSP_005031.
      L -> LL (in MD-BBS).
      /FTid=VAR_011336.
      Missing (in MD-BBS).
      /FTid=VAR_011337.
      R -> W (in BBS1).
      /FTid=VAR_015817.

Query Match 3.1%; Score 294; DB 1; Length 4684;
Best Local Similarity 19.7%; Pred. No. 6.5e-08;
Matches 326; Conservative 283; Mismatches 607; Indels 438; Gaps 71;

QY 249 VHATHTRALACIHE-----EEGSPRTP-TKSAIHLKRDIKLVGSLPKYVSESMKDL 302
DB 611 VYRLHERLVAITEYNLRKAGVAPATQVTLQSVQRPELSDTLR-----YLQDLL 666
QY 303 LLSPTQKVAIQGS-----MPLNEM--KIRSTEQNRYYQEQQIQIKADKLNSSSSSEASL 355
DB 667 AWVEENQHRVDGAEGVDLPVSAQLGSHRGLQS--IEEQAKTERRARSDGQLSPATR 724
QY 356 AQOQELFNHSELIAOSSRFNLHEVGRKSVKGVSPVNHKRSHEL-----SPSDAPSNE 411
DB 725 GAYRDCI-----GRDLQYAKLLNSKARLSRLSLHSFVAAATKEL 766
QY 412 LYRNETVAISPCKORVEDTTLPRGAAPANASARSSSAHAMF--HAQSKFKLAQTWSL 468
DB 767 MWLINE-----KEEEVGFDWDSDNTNMTAKESYSALMRELEKKEKIKELQN--- 814
QY 469 MKKPATPRKVR-DTSIQPSVKLYDSELYMOTCINPDPPFAATTIDPFLASTWYLDQAVD 527
DB 815 ---AGDRLLREDHPARPVTSFOALQTC----- 840
QY 528 RHQADFKKWLNALVSPADLDLANNKIDVGLFNEVRNKB-----LVVAPTKEEQSNNY 582
DB 841 -----W-SMWLQCCCEIAHLKENAAYFOFSDVREAGQLQKQEAFLRRKNSCDRS 891
QY 583 LTKYRLTLRKAARELVFFSEQMRLPSCVAVVYVVKQALRIRSDRL-----HLDVVMQRTIL 639
DB 892 ATVTRLEDLQDAQD-----EKQQL--NEYKGLHSLGLAKAKAVVQLKPRHPAHPMRGLP 945
QY 640 ELLLCENPLNRLGLVNVFGEKIQMSNRDVIQGLSTFINRLFRNKCEQRYSKAYTLTE 699
DB 946 LLVAVCD---YKQVEVTVHKGDQC-----LVG-----PAQPSHWKVLSS 981
QY 700 EYAEITKKHSLKILFLPFLDQAKQRIK-----HNPCLFVKKSPHKEKIDIL-----LR 751
DB 982 SGSEA---AVPSVCLVPPNPQVEQVAVTRLEAQHQLVTLHQLHVDMSLLAQWSLR 1037
QY 752 FSEELLANTGDTT-----RELRLGLVLO-HRQFTL-DEFD----- 785
DB 1038 RDVQIRSLATFTLTKPEEQRALHSLHLYQAFRLDSQDAGFGPEDRLMAERYGS 1097
QY 786 --YAFNLAVDLRDQ-----VRL-----TRVVEVILLRDDLRLQLRVP 821
DB 1098 CSHHYQQLLOSLEQAQESRCORCISELKDRLQLEACEITVRLRLPLD-----KEP 1152
QY 822 AISRLQIRFNVKLA-----LGALEGEANFOLGGDIAAQDIDVGHREKTLISLLW 868

```

```

DB 1153 ARECAQRIABEQKQAQAEVGLGKGVARLSABAKVUALPEPSPAAPTILSELELTGLKLE 1212
QY 869 QLIYKFRSPKFAAATVLOKWRHRHVLHVVIQRIIRHKELMRH---RAATVIOAVFRG 924
DB 1213 QV-----RS-----LSAIVLEK--LKTSIVIRGTQGAEEVLRAHEBQLKEAQPAT--- 1258
QY 925 HQWKYVKLPKFTERTOAIILOKFTTRYLAOKLYOSIHSIIITIQWRWAQQLG---RQ 980
DB 1259 -----LPELEATKAS--LKRLAQAEAQOFTFDALRDEL---GAOEYGEELQOR 1303
QY 981 HRQFVLELRAAIFLQIRMR---LFAKKLLAAAEFARLQSKOKQOAAASYTQMWQRTY 1037
DB 1304 HGERDVEVER-----WEERVAQLLERWQVLAQTDVQRELEQ-----LGRQLRY 1349
QY 1038 Q-----LGRIOREHFLRQDL-----IMFVORMRSKWSMLEQ-----RKEFQQLK 1078
DB 1350 RESADPLGAWLODARRRQEQIQAMPLADSOAVREQLRQEQALLEIERHGEKVEECOREFA 1409
QY 1079 RAAINIQWRRAKLSMEKCNADYLALRSSVLKQAVRKATIQMRID-RNHYSYL---RK 1133
DB 1410 KQYINAKDYELQLVYTKAQLEBPVAPKPKVQSGSESVIQYVDLRTYSELTLTSQ 1469
QY 1134 NVICLOQLR-----AIMKREORENVYLRNASILVQKYRMQOMIQ-----DRN 1180
DB 1470 YIKFISETLRMBEERLAEQQAERE---RLAEVEAALEKQRLAEAAHAQAKAQAERE 1526
QY 1181 AYLRTRKCIINVORWRATLOMERERKNYLHLTTTKRIQIKPRAKREMKKQPAEFLQLK 1240
DB 1527 AKELQQRIQEVVRBEAAVDAQOQKRS---IQEELQQLRQSEABIQAKARQAEE--- 1581
QY 1241 KVTILVQKRRALLQMKERQBYLHLREVTIKIQ---BRFHAQKSMRPMRAKYRTOAA 1296
DB 1582 -----RSRLRIBEE-----IRVVLEQLAEATERQGGAEGLQALRA--RAEEAE 1623
QY 1297 VSCLOMHWNRHILAKERNPSFLQIRQAATLQRYRAB-----LNMIKOLKSYAQ 1346
DB 1624 AQKQQAQAEARLRQVQDESQRKQAEVELASRVKAEABAEAREKQALQALEELQAE 1683
QY 1347 -----LQQAATITQTRYAKKA--MQQWVLYQK-----QREAITKVQR 1383
DB 1684 EAEWNLQAEVERARQVQVALETAQSAEAELOSKEASFABKTAQLERSIQEHHVAQL 1743
QY 1384 RYGNLEMRKQIIVYQKQOAVIRLOKWNRSIRD-MRLCKAGYRRIRLSLS----- 1434
DB 1744 REAEARRAQQAERAEAREERQLERWQLKANEALRLRLQAEVLQKSLAQAEAEKOK 1803
QY 1435 -----IQKWRATVQARQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473
DB 1804 EEAERARRRGRKAEQAVRQRELEABQELEKQORLASGTACQRLAAEQELIRLRAETQGE 1863
QY 1474 QORREPE-----MKRBAVVQIRFRPARCAMLKARQDYQLIOSSVILVQKFR---ANRSM 1526
DB 1864 QORQLLEELARLQREAAATQKQLEABLAQVRAEMEVLLKASKAKAEESRSTSEKSK 1923
QY 1527 KQARQEVQVLTAVHLQKFKRKLMIQRNCFOLLRCMSPGFORARGFARK----- 1581
DB 1924 ORLEASAGRPRELA---EEAARLRAEAEAKRQRLAEBDAARQARAEAEVLAELKLAATG 1980
QY 1582 RFQALMTPEMMDLIRQKRAKVIRYWRGVLIRRR---QKHQGLDIDIRKLAQLROEAK 1637
DB 1981 EATRLKTEABIALKEAEANERLRLAEBDEAFQORRELEBQAQHKADIERLAQLR---K 2037
QY 1638 AVNS-----VRCKVQEAVERFLRGF 1657
DB 2038 ASDSELERQKGLVEDTLRQRRQVEEILAKASF 2071

RESULT 9
PLEI CRIGR
ID PLEI_CRIGR
AC Q90I55,
DT 16-OCT-2001 (Rel. 40, Created)

```

16-OCT-2001 (Rel. 40, last sequence update)  
 15-MAR-2004 (Rel. 43, last annotation update)  
 Plectrin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated protein) (IFAP300) (Fragment).  
 PLEC1.  
 Cricetus griseus (Chinese hamster).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.  
 NCBI TaxID=10029;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20334248; PubMed=10873583;  
 RA Clubb B.H., Chou Y.-H., Herrmann H., Svitkina T.M., Borisy G.G., Goldman R.D.;  
 RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a hamster plectrin ortholog.";  
 RL Biochem. Biophys. Res. Commun. 273:183-187(2000).  
 RN [2]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=96215219; PubMed=8626512;  
 RA Malecz N., Folsner R., Stadler C., Wiche G.;  
 RT "Identification of plectrin as a substrate of p34cdc2 kinase and mapping of a single phosphorylation site.";  
 RL J. Biol. Chem. 271:8203-8208(1996).  
 CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and microfilaments and anchors intermediate filaments to desmosomes or hemidesmosomes. May be involved not only in the crosslinking and stabilization of cytoskeletal intermediate filaments network, but also in the regulation of their dynamics.  
 CC -!- SUBUNIT: Homodimer or homotetramer.  
 CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N- and the C-terminus can bind integrin beta-4.  
 CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM INTERMEDIATE FILAMENTS DURING MITOSIS.  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 32 plectrin repeats.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 EMBL; AF260753; AAF70372.1; -.  
 DR HSP; Q01082; 1BKR.  
 DR InterPro; IPR001589; Actbind actinin.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR001101; Plectrin\_repeat.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectrin; 20.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00250; PLEC; 34.  
 DR PROSITE; PS00019; ACTININ\_1; PARTIAL.  
 DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding; Phosphorylation.  
 FT NON TER 1  
 FT DOMAIN 1259  
 FT DOMAIN 1260 2544  
 FT DOMAIN 2545 4473  
 FT DOMAIN 4474 4473  
 FT ACTIN-BINDING.  
 FT CH 1. 74  
 FT DOMAIN 87 189  
 FT CH 2.  
 FT SPECTRIN 1. 508  
 FT REPEAT 449 508

FT REPEAT 529 613 SPECTRIN 2.  
 REPEAT 626 719 SPECTRIN 3.  
 FT REPEAT 1104 1204 SPECTRIN 4.  
 FT DOMAIN 1258 2548 COILED COIL (POTENTIAL).  
 FT REPEAT 2615 2652 PLECTIN 1.  
 FT REPEAT 2653 2690 PLECTIN 2.  
 FT REPEAT 2691 2728 PLECTIN 3.  
 FT REPEAT 2729 2766 PLECTIN 4.  
 FT REPEAT 2770 2804 PLECTIN 5.  
 FT REPEAT 2905 2942 PLECTIN 6.  
 FT REPEAT 2943 2980 PLECTIN 7.  
 FT REPEAT 2981 3018 PLECTIN 8.  
 FT REPEAT 3019 3056 PLECTIN 9.  
 FT REPEAT 3057 3094 PLECTIN 10.  
 FT REPEAT 3274 3311 PLECTIN 11.  
 FT REPEAT 3312 3349 PLECTIN 12.  
 FT REPEAT 3350 3387 PLECTIN 13.  
 FT REPEAT 3388 3425 PLECTIN 14.  
 FT REPEAT 3429 3463 PLECTIN 15.  
 FT REPEAT 3609 3646 PLECTIN 16.  
 FT REPEAT 3647 3684 PLECTIN 17.  
 FT REPEAT 3685 3722 PLECTIN 18.  
 FT REPEAT 3723 3760 PLECTIN 19.  
 FT REPEAT 3764 3797 PLECTIN 20.  
 FT REPEAT 3800 3834 PLECTIN 21.  
 FT REPEAT 3852 3889 PLECTIN 22.  
 FT REPEAT 3890 3927 PLECTIN 23.  
 FT REPEAT 3928 3965 PLECTIN 24.  
 FT REPEAT 3966 4003 PLECTIN 25.  
 FT REPEAT 4007 4041 PLECTIN 26.  
 FT REPEAT 4043 4094 PLECTIN 27.  
 FT REPEAT 4197 4234 PLECTIN 28.  
 FT REPEAT 4235 4272 PLECTIN 29.  
 FT REPEAT 4273 4310 PLECTIN 30.  
 FT REPEAT 4311 4348 PLECTIN 31.  
 FT REPEAT 4349 4386 PLECTIN 32.  
 FT DOMAIN 4039 4089 BINDING TO INTERMEDIATE FILAMENTS (BY SIMILARITY).  
 FT DOMAIN 4414 4429 4 X 4 AA TANDEM REPEATS OF G-S-R-X.  
 FT MOD\_RES 4328 4328 PHOSPHORYLATION (BY CDC2).  
 FT SEQUENCE 4473 AA; 509015 MW; E144615D361E3484 CRC64;  
 Query Match 3.0%; Score 287; DB 1; Length 4473;  
 Best Local Similarity 19.2%; Pred. No. 1.5e-07;  
 Matches 358; Conservative 317; Mismatches 736; Indels 456; Gaps 77;  
 QY 161 ETRGKQATIP--PDNLAAP-----TPTLKNVKSANDMRPRITDDLEDQATNKT 213  
 DB 960 EGIKGVARLSAEAKVLAPEPSAAPTIRSELE-----LTGLKLEQVRSLSAI 1009  
 QY 214 FVVKHSETINISLDTDCSRIDGQPHTPINKTTTIVHATHTR--ALACIHEEGSPSPR 270  
 DB 1010 Y-LEKLTISIVIRSTQGAEEVLKAHERQLKEAQAVPATLQELEATKASKLKRAQAQAQ 1068  
 QY 271 TPTKSAIHLKDKIKLVGSP--RYKSESKMKSLSLSPOTKYAIQSGMNMNMKIRSIIE- 328  
 DB 1069 QPVFDTLRDELRCAGVEGVRGRLQQRGGERDVEVERWRKRVNQLERWQAVLAQIDVQREL 1128  
 QY 329 -----QNRVYQF-----QQIQIKAKDLNSSSSSEASIAQQQ---EFLVHS 366  
 DB 1129 EQLGRLRYRESADPLSLWLODAKRRQEQIQAVPIPSQAAREQJRKQKALLEIERHG 1189  
 QY 367 ETLAOSRFNLEHVGKSVK-----PKQVRVED--TTLPRSAAPANASARSSSAHA 451  
 DB 1189 EKVEECQFAKQYI--NAIKDYELQVTVYKAQLEFVSPAKPKVQSGSESVIQEVDLR 1246  
 QY 411 SLRYNETVAISP-----PKQVRVED--TTLPRSAAPANASARSSSAHA 451  
 DB 1247 TRYSELTTTTSQYIKFISETLRMEERERLAEQQAAREERLAEQQAAREERLAEQQAARE 1306  
 QY 452 WPHAQSKFKLAQTMKMKPATPRKVRDTSFQPSKLYDSEL--YMQTCINPDPAATTT 510  
 DB 1307 QAKQAQ-ELEAQELQRMQEEVAREEAAVNAVQAQQKRSIQBELQHROSRAEIQAKAQ 1365

```
QY 511 IDPFLASTWYLD-----QAVDRHQADFKKWLNALVSIPLADLADLNN-KIDVGL 560
Db 1366 VEAABSRMKIEEIRVVRLQLETTERRQGGAEGLQALRAAEAEAKRQAQAEERL 1425
QY 561 FNEVRNK-----ELVV-----APKBEQSNVYTKVRIET-----LRKAAYEL 598
Db 1426 RQVODESQRKQAEABTLRVKAQAEAEQKORALQALEELRLQAEAEERLRQAQER 1495
QY 599 FFEQMRPLPC-----SKVAVVYNKOALRIRSDRNHLHDVVQMRTILLLCFNPL 648
Db 1486 ARQVQVLAETAQSAEVELQSKRSPAEKTAQLETLQBEHVTVQLREKAE-----1537
QY 649 WRLGLGVVFGKIQMOSNRDIVGLSTFILNRLFNKCEQRYSKAYTLTBYEATIKKH 708
Db 1538 --RRAQQQAEARARAEAREEL-----ERWQLKANEALRLRLQAEVAQOR 1581
QY 709 SIQKILFLPLFDQAKOKRIKVHNPCLFVKKSPHKETKDILLRFSSELLANIGDITREL 768
Db 1582 SLAQ-----ADAEKQKE-----EAEREAR 1600
QY 769 RIGVYLQH--ROTFLDEPDYAFNNLAVLDGVRITRVVVEVILLR-----DDLTRQLRVP 821
Db 1601 RRGKABEQAVRQELAEQLEKORQALAECTAQORLAAEQELIRLRAETEQGQORQLLEE 1660
QY 822 AISRLQR-----IFNVKIALGALGEANFQGGDIAODIVDGHREKTLISLLWQLI 871
Db 1661 ELARQREATAHKKQOELEAEIA-KVRAEMEVILASKARAEESRSTSEKSKQRLBABA 1719
QY 872 YKFRSPKPHAA-ATVILQKWRRHLHVITQIRIRHKELMRHRAATVIAQVFRGHQMRKY 930
Db 1720 DRPFRLAEAEARLALAEAKR-----QRQLAEEADAARQAEAEVLT-----EKL 1765
QY 931 VKLFKTER--TQAAIILQFTR-----RYLAQKQLYQSVHSIITQIRWRAQQLGRQR- 982
Db 1766 AAISEATRIKTEABTALKEKAENRERLRLAEDEAFQRR-----RUEEQALHKA 1815
QY 983 ---QREVELREAA-----IFLQRIWRRLFLAKLL-----AAAEFARIQ-----1018
Db 1816 DIEERLAQLKASESELERQKGLVEDTLQRRQVEEILALAKVSFEKAAAGKAELELEIG 1875
QY 1019 --RSQKQAAASYIQMOWTYQLGRIOHFEFLRQDLIMFVQRRWRSKWSMLQORK-----1072
Db 1876 RIRSSAEDTWRSEKQAEQAAQRQQLAAEEQRRREABERVQKSLAAEEAEARQKAALE 1935
QY 1073 EFQOLKRAAINIQORWRKLSMEKONADYALRLSSVLKVQVYKATIQMRIDENHYYSIR 1132
Db 1936 EVERLK-AKVEARLRERAEQESARQLOLQAEAAKRIQAEKA-----1979
QY 1133 KNVICLQRLRAIMKMQRENYL-FLRN-----ASILVQKRYMRQMIQDRNAVYLRTRK 1187
Db 1980 -HAFVVQREELQQTLOEQSMLERLGEAEAAARAAAEAEAREQ--AEAEA-AQSRK 2035
QY 1188 CIINVR-----RWATLQMRERKNYLHQTITKIQIKFRANKMKQKQAEFLQ 1238
Db 2036 QVEEERLKQSEBQAEQAAQAAAEKURKEABEABARRAAQAEQAALQKQAAAEWK 2095
QY 1239 LKKVTLVQRRRALIQMERQERYHLREVTIKLQRRPFAQKSM---RFWEAKRYGTQA 1295
Db 2096 HKKFA-----EQTLRQKAQVQELTTLA--LQLEETDH-QKSILDBELQRLKAEVTEA 2145
QY 1296 AVSCIQ-----MEWRHILRKRENS-----FLQIRQAAITLQRRY 1331
Db 2146 ARQSQVBEELFSRVQMEELGKIKARIEAENRALILDKONTQRFLEAEFKMQVAABE 2205
QY 1332 RARLNMIKQLSKYAQLKQAA--ITTIQTRVRAKAMQKVLYQ-----KOREAIIKVORR 1384
Db 2206 AARLSVAQ--EAARLQLAEDLAQORALAEKMKKQVQAEATRLKAEAILQOQKE 2263
QY 1385 YRGNLEMRKQIEVYQORQAVIRLQKWRSI-----RDMRLCKAGVR-RIRLSLSIORK 1438
Db 2264 LAEQARRLIQEDKEQMAQOQLVETQGTQFLEVERQORLEMSAEABRLKLRMAEWS-RAQ 2322
```

```
QY 1439 WRATVQRRRRE-----IFLSTIRKVRLMQAFIRATILMRQORREFFEMKR-RAA 1486
Db 2323 ARABEDAQRFKQAEIEGKELHETELATQEKVTLVQ-----TLEIQOQSDHDAELREA 2377
QY 1487 VTIQRRFR-----ARCAMLKARQDQLOISSLVIL-----VQKFRANRSMKQARQEFVQ 1535
Db 2378 IAELEREKELKQBAKLLQKLS--EEMQTVQOEQLQETOALQKSLFKSDKSLLOREFFIE 2436
QY 1536 LRTIAVHLQOKFRGKIRIMIEQRNCFQLLRCSMPFGFQARAGFMRKBFQALMTPEMDLI 1595
Db 2437 -----QEKAKLEQLFQDEVAQAQQLR-----EEQQRQQRQMEQKQELV 2475
QY 1596 -----RQKRAKVIQRYWR--GYLIRERQKHQGL-----DIKRIALQPOEAKAV- 1639
Db 2476 ASMEEARROCEAEHAEVRRKQEBLQHLQROQOEKILAEENQRLERLQRLLEEHRAL 2535
QY 1640 ---NSVRCKVQEAQVRFLL-RGRFTASDALAVLSQ-----LDRLSRTVPHLLMWCSEFMS 1688
Db 2536 AHSEEIATAQAAAALPNGR---DALDQPSMEVPEHAFAEGLRQKVPAL-----2582
QY 1689 TFCYGINMAQIRSEVDKQILIERCSRILNLARVNSTVTTFQEGGLVTIAQMLLRWCDKD 1748
Db 2583 ----QUREAGILSABEQLRLEQGHITVAELSQRED--VRQYLQ-GRSSIAGLLKPTDEK 2635
QY 1749 SEIPNTL 1755
Db 2636 LSVVTAL 2642
```

RESULT 10  
HMW2 MYCGE ID HMW2 MYCGE STANDARD; PRT; 1805 AA.  
AC P47450;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, last sequence update)  
DT 16-OCT-2001 (Rel. 40, last annotation update)  
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory protein 2).  
DE HMW2 OR MG218.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Nguyen D.T., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L., Tomb J.-F., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Peterson S.N., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.; "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403 (1995).  
RP SEQUENCE OF 557-659 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random sequencing.";  
RL J. Bacteriol. 175:7918-7930 (1993).  
CC -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY SIMILARITY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way







RA Rodriguez O.C., Cheney R.E.;  
RT "Human myosin-Vc is a novel class V myosin expressed in epithelial  
RL cells.";  
RL J. Cell Sci. 115:991-1004(2002).  
CC -!- FUNCTION: May be involved in transferrin trafficking. Likely to  
CC power actin-based membrane trafficking in many physiologically  
CC crucial tissues.  
CC -!- TISSUE SPECIFICITY: Expressed chiefly in non-neuronal tissues.  
CC Particularly abundant in epithelial and glandular tissues  
CC including pancreas, prostate, mammary, stomach, colon and lung.  
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -!- SIMILARITY: Contains 1 dilute domain.  
CC -!- SIMILARITY: Contains 6 IQ domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF272390; AAF78783.1; .  
CC HSP; P08799; 1VOM.  
CC Genew; HGNC:7604; MYO5C.  
CC InterPro; IPR002710; DIL.  
CC InterPro; IPR000048; IQ\_region.  
CC InterPro; IPR001609; myosin\_head.  
CC Pfam; PF01843; DIL; 1.  
CC Pfam; PF00612; IQ; 5.  
CC PRINTS; PR00193; myosin\_head; 1.  
CC PROSITE; PS00063; myosin\_head; 1.  
CC ProDom; PD003376; DIL; 1.  
CC ProDom; PD000355; myosin\_head; 1.  
CC SMART; SM00015; IQ; 5.  
CC SMART; SM00242; MYSC; 1.  
CC PROSITE; PS00096; IQ; 4.  
CC MYOSIN; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Coiled coil; Polymorphism.  
FT DOMAIN 1 755 MYOSIN HEAD-LIKE.  
FT DOMAIN 756 779 IQ 1.  
FT DOMAIN 780 806 IQ 2.  
FT DOMAIN 807 829 IQ 3.  
FT DOMAIN 830 854 IQ 4.  
FT DOMAIN 855 884 IQ 5.  
FT DOMAIN 884 1351 COILED COIL (POTENTIAL).  
FT DOMAIN 1574 1679 DILUTE.  
FT VARIANT 522 522 P->L.  
FT VARIANT 634 634 /FTID=VAR\_010646.  
FT VARIANT 634 634 L->S.  
FT /FTID=VAR\_010647.  
SQ SEQUENCE 1742 AA; 202793 MW; 25DD3082A7EB5AAB CRC64;  
Query Match 3.0%; Score 281; DB 1; Length 1742;  
Best Local Similarity 17.7%; Pred. No. 8.9e-08;  
Matches 291; Conservative 274; Mismatches 515; Indels 560; Gaps 77;  
QY 415 NETVAISPPKQKQVETTLPRSAAPANASARSSAHAWPHAQSKFKLAQTMSLMKKEAT 474  
Dd 382 SETV-----VKEMTPQAVNARDALAKKIYAHLPDFIVE---RINQALQFSGKQHT 429  
QY 475 PKVRDTSTQPSVKLYDSLYMWTCTIN----- 501  
Dd 430 FTGVLDIY---GFETPDVNSFQFCINYANEKLQQQFNMHVFKLEQSEYMKEDIPWTLID 486  
QY 502 -PDFFAATITDPFLASTMYLDEQAVDRHQADFKKWLNALVSIADLADLNKIDVGKL 560  
Dd 487 FYNQFVIDILIAKMGILDECLPLHGT-ENWPQL-----YNNFVNRNPL 535  
QY 561 FNEVR--NKLIV---APTEKQSGMYLTKYR-----LETIRKADEL---FFSEQMR 605  
Dd 536 FKPRMSNTSFVIQHFADKVEYKCEGFLKRNRTVYDMLVELTRASKPHLCANFFQENPT 595

QY 606 LPSCKVAVYVYNQALRIRSDRNHLHDVVMORTILELLCENPLWRLGLGVFGEKIQMG 665  
Dd 596 PP-----SP-----FGSMITVK 607  
QY 666 SNRDIVGLSTFILNRLFRNKCEQRYSKAYTL-----TEEYAEITKKHSQKILFLP 718  
Dd 608 SAKQVIKPN---SKHPTTVGSKFRSSLYLLMETLNATTPHYVRCIKPNDEK---LP 658  
QY 719 FLDAQOKRIVKUNPCLFVKKSPHKETKDILLFSSSELLANIGDITRELRLRGLVYLOHRQ 778  
Dd 659 F--EFDKSRIVQQLRACGVLET-----IRISAQSYPS-----RW 690  
QY 779 TFLDEPDYAFNNLAVDLRDGVRLTR-----VVEVILLRDLTRQLRVPALSRLQR 828  
Dd 691 TYIEFIS-----RYGILMTKQELSFSDKKEVKVVLHR----- 723  
QY 829 IFNVKLALGALGAN-PQLGGDIAAQIDVDGHRKTLSLMLQIY--KFRSPKFHAATV 885  
Dd 724 -----LIQDSNQYQFG-----KTKIFFRAGQVAYLEKRLDLRQSCVM 762  
QY 886 LOKWRRRHHLHVLIQRIIRHKELMRHRAATVIOAVPRGHO-MRKYVKLFTKERTQAII 944  
Dd 763 VQR-HMRGWL-----QRKKFLRERRAALIQQYFRGQQTVRKAITAVALKEAWAII 813  
QY 945 LQKFTRYLAQKQIYQSYH-SIITIQRMWRA-----QQLGRQH-----R 982  
Dd 814 IQHCRGYLV-RSLYQLIRMATITMQAYSRGFLARRRYKMLEKHAVALIKQYARAWLAR 872  
QY 983 QRFVELREAAFLQRIWRRLFAKL-----LAAETARLQSQQAAA 1026  
Dd 873 RRFQSIIRFVLIQLTATYRVQRLQKLEQDNKENHGLVEKLTSLAALRAGVEIKQLEAF 932  
QY 1027 ASYIQMOWRTYQ-----LGRIORH-EFLQRORDLIMFVQRRMRKSWMLEQR 1071  
Dd 933 LEKATATRRNYEKGKRYRDAVEKLAQKLNSELETKEQEI--QLKQETELKEK 989  
QY 1072 KBFOOLKRAAINTQORWRKLSMRKNADYALRSLVVKQAY-----R 1115  
Dd 990 MD-NLTKQLFDVQKEERQRMLEK-----SPELKTQDYEKQTSIKBIKAKDE 1039  
QY 1116 KATIOMRIDRNHYYS--LRKNVICLOQLRAIMKREORENLYRLNASLTVOK----R 1168  
Dd 1040 KMQLQHLVEGHEHTSDGLKAEVAVLSKVQKVTISEFKEIE--LLQAKIDVSKHVQSQK 1096  
QY 1169 YRMEQOM-----IQDRNAVLTETKCIINVQ-----RRWEATLQMRERKKNYL 1210  
Dd 1097 REMEEKMSBITKQLLESYDIED---VRSLSVEDLEHNLNEDGELMFA-----YE 1142  
QY 1211 HLQTTTTRIQIKFRKREMKKQRAEFLOLAKKTVLVQKRRALLQMRKEQVYHLREVT 1270  
Dd 1143 GLKKAATVLESHTFQSQDKCYEKEIEALNFKVVHL-----SQEINHLQKL- 1186  
QY 1271 IKLQRRFHAOKSMRFRKRYRGTOAAVSCLOMHWNRHLLKRRNSFLQLRQAATLQRR 1330  
Dd 1187 --FREENDINESIHETVRLTSENMMIPDFKQIIE--LEKQKDLIELRNEQA---EK 1238  
QY 1331 YRAPLNMIKOLSKVAQKQAAITQTRYAKKAMQKVLYQKORAIL-KVQRRYRGNL 1389  
Dd 1239 MKGKLELSNQLRHSREECT-----ORKALEAQNEIHTKEKELDKIQEMQEASD 1290  
QY 1390 EMRKOIIEVYQKQ--RQAVIRLQKWRISIR--DM--RLCKAGYRRIRLSLSI----- 1435  
Dd 1291 HLKQAFTESEVKNFQEAASRLTLENRDLREELDMKORVKKLQOVKTLSTIKGAND 1350  
QY 1436 -----QRKWRATVQARRQREI-----FLSTI 1456  
Dd 1351 VHSSSGPEKYLGMQYKREDEAKLIQMLILDLPKRGVWVNMIPGLPAHILFMCVRYADSL 1410  
QY 1457 RKVELMQAFIRATL-----LMRQORREFFEMKRAAVIQRRFRRCAMKARQDYQIQS 1511  
Dd 1411 NDANMLKSLMNSTINGIKVQVKEHLEDFEM-----LSFWLNTCHFINCLKQYS---- 1459  
QY 1512 SVILVQRKFRANRSMKQAR-----QEFVQ-LRTIAVHLQOKFRGKRLMIFQRCFQL 1562



```

Db      948 KHK-----EELLAQQAELQELQEQATALASQODTVQAKDLARQOHELELQON 999
QY      1135 VICLOQRRAIMKREORENYELRNASILVQKRYR--MRQOMTQDRNAVLTREKCIIN 1191
Db      1000 --AFNQASISLNKQRBOLTNQVKVLHGEL--KKREKLTLDRLAAREKQHKDAEIN 1055
QY      1192 VQRWRATLOMRERKNYVHLQTTYR-IQIFRAKREMKQRA-----EFLQL 1239
Db      1056 -QR-----FKQFENEYADFQAKKRELQELNOTLRNLEQNASLKKRNOLTLDFALL 1107
QY      1240 KVVTLVQRRALLQMREROSYHLREVTTIKQRPFAKQSMRPMKRYGTO----- 1294
Db      1108 RKVQHTQNR---VQNTQIKFELKKNQKQFQASDERALQALILKLRSPASKLQIQR 1164
QY      1295 --AAVSCLOMWRNHLKRERNFLOLQAAITLQRYRABLANMIKOLKSYAQ----- 1346
Db      1165 EALAIQKLEFDRDEQKQSEINNAKIQLEQFKLEKQNFQAKQKQLEFQKQRLDVEK 1224
QY      1347 --LKQAAITOTRYRAKKAMQOVLYQKQREAILIKVORRYGNLEMKQIIE----- 1396
Db      1225 RLKQKVLQKLSKSYLYTKRADLSQQQ-----LOHYANLLEKKEKLTAKRALDK 1278
QY      1397 ----VYQORQAVIRLOKWRISIRDMRLCKAGYRIRLSSLSIOR-----KWRATVQARRQ 1448
Db      1279 KHRAYGKMAQFVSEL-----ROEKQLLSAQKQVDDKSRLLQKQNRH 1321
QY      1449 REIFLSTIRKVRMLQAFIRATLIMROQRREPEMKRAAVIQRFRACAMLKARQDYQL 1508
Db      1322 LQNLSETKKKKQS-----LEHDINKFQRRKEAV----- 1351
QY      1509 IOSSVILVORKFRANRSMQARQEFVQLRTIAVHLQKQKPRKRLMEQRCNCPQLLRCSMP 1568
Db      1352 --SSIL-----NSHKKLQKXEGELQIL-----QKLSLKKTLQEQE--PSKL----- 1389
QY      1569 GQPARARGMARKRFOALMTPENMDLIQKRAKVIQRYWRGYLIRRRQKHQGLDI--- 1625
Db      1390 -YQREK-----LDQRTLSKLHRELKAQNATAFKKNREVLIEYNY 1430
QY      1626 -RKRIQLRQEAQKAVNSVRCKQEAVFRLGRFIASDA--LAVLSQDLRLSRTVPHLMM 1682
Db      1431 YKKEQLRLTEKSEFQNNKRLFEYFKRIENIEKKEAHKIVLEBQTKR-----HL--- 1483
QY      1683 CSEFMSTFCYGINMAQAIRSEVDKOLI-----ECSRIILNLARYNSTVTNTPQE 1731
Db      1484 -----VETEAVKLHLQKQSIISKQELKEIKERVSRDISHNTKQREELNSLLHQ 1532
QY      1732 GGIIV--TIAQMLLRWCDDKSEIWTNLTLLIWPFAHCPKRRKIIH--DYMTNPEATVMYRE 1787
Db      1533 NKLLQKNLAEEREIENKNSLITQKIQTAKQLSE--KEARILKLEKMRVAVEQYQQA-E 1589
QY      1788 TKKLVARKEKMKONARKP--PQMTSGRYKSOKINFPTCPLSLPDPF 1832
Db      1590 ITELKTRNADLEKNDKNHLPPLF-----KINGNDMNPYPYPWF 1629

RESULT 13
TRHY HUMAN
ID TRHY HUMAN STANDARD; PRT; 1898 AA.
AC Q07283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Trichohyalin.
GN TH OR TRHY OR THL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93280194; PubMed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,

```

```

RA      Steinert P.M.;
RT      "The structure of human trichohyalin. Potential multiple roles as a
RT      functional EF-hand-like calcium-binding protein, a cornified cell
RT      envelope precursor, and an intermediate filament-associated (cross-
RT      linking) protein."
RL      J. Biol. Chem. 268:12164-12176(1993).
RN      [2]
RP      SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX      MEDLINE=93315897; PubMed=7686953;
RA      O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT      "Trichohyalin: a structural protein of hair, tongue, nail, and
RT      epidermis."
RL      J. Invest. Dermatol. 101:65S-71S(1993).
CC      1- FUNCTION: Intermediate filament-associated protein that associates
CC      in regular arrays with keratin intermediate filaments (KIF) of the
CC      inner root sheath cells of the hair follicle and the granular
CC      layer of the epidermis. It later becomes cross-linked to KIF by
CC      isodipeptide bonds. It may serve as scaffold protein, together
CC      with involucrin, in the organization of the cell envelope or even
CC      anchor the cell envelope to the KIF network. It may be involved in
CC      its own calcium-dependent postsynthetic processing during terminal
CC      differentiation.
CC      2- SUBUNIT: Monomer (Probable).
CC      3- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC      the inner root sheath (IRS) of hair follicles and medulla, and in
CC      the filiform papillae of dorsal tongue epithelium (Probable).
CC      4- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC      the epidermis.
CC      5- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC      calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC      entirely alpha-helical, configured as a series of peptide repeats
CC      of varying regularity, and are thought to form a single-stranded
CC      alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC      the most regular and may bind KIF directly by ionic interactions.
CC      Domains 5 and 7 are less well organized and may induce folds in
CC      the molecule. Domain 9 contains the C-terminus, conserved among
CC      different species.
CC      6- PTM: Substrate of transglutaminase. Some 200 arginines are
CC      probably converted to citrullines by peptidylarginine deiminase.
CC      7- SIMILARITY: In the N-terminal section; belongs to the S-100
CC      family.
CC      8- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; L09190; AAA65582.1; -.
DR      PIR; A45973; A45973.
DR      HSSP; P02633; 4ICB.
DR      Genew; HGNC:11791; THH.
DR      MIM; 190370; -.
DR      GO; GO:0005856; C:cytoskeleton; NAS.
DR      GO; GO:0005509; F:calcium ion binding; TAS.
DR      InterPro; IPR001751; CaBP_S100.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR002017; Spectrin.
DR      Pfam; PF00036; ehand; 1.
DR      Pfam; PF01023; S_100; 1.
DR      ProDom; PD003407; CaBP_S100; 1.
DR      PROSITE; PS00018; EF_HAND; 1.
DR      PROSITE; PS00303; S100_CaBP; 1.
DR      Keratinization; Calcium-binding; Repeat; Citrullination.
KW      DOMAIN 1 91 S-100 LIKE.
FT      CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT      CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT      DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF
FT      R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT      REPEAT 314 326 1-1 (APPROXIMATE).

```

```

FT REPEAT 327 339 1-2 (APPROXIMATE)
FT REPEAT 340 351 1-3 (APPROXIMATE)
FT REPEAT 352 364 1-4
FT REPEAT 365 377 1-5
FT REPEAT 378 390 1-6
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L
FT REPEAT 391 444 2-1
FT REPEAT 397 402 2-2
FT REPEAT 403 408 2-3
FT REPEAT 409 414 2-4
FT REPEAT 415 420 2-5
FT REPEAT 421 426 2-6
FT REPEAT 427 432 2-7
FT REPEAT 433 438 2-8
FT REPEAT 439 444 2-9
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1
FT REPEAT 953 982 4-2
FT REPEAT 983 1012 4-3
FT REPEAT 1013 1042 4-4
FT REPEAT 1043 1072 4-5
FT REPEAT 1073 1102 4-6
FT REPEAT 1103 1132 4-7
FT REPEAT 1133 1162 4-8
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1752 1752 F -> L (IN REF. 2)
FT CONFLICT 1794 1801 QERDQYR -> RSETGSTG (IN REF. 2)
FT CONFLICT 1857 1857 Q -> K (IN REF. 2)
FT CONFLICT 1880 1880 V -> G (IN REF. 2)
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947F562E31D CRC64;

Query Match
Best Local Similarity 2.9%; Score 277; DB 1; Length 1898;
Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

QY 891 RRHLVHVIVQRIIRHKLMMRRHRAATVIOAVFRGHQMKVYKLFKTBRTQAAIIQKFTFR 950
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 648 RHLLKSEGEERHRLRQERQERREORLKRREERLEORLKRHE-----ERRE 701
QY 951 RYLAQKQLYQSYHSIT-IQW-WR-----AQQLG----- 978
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 702 QELABEEQEARERIKSRPKWQWLESEADARQSVLEAPQAGRAEAPQEQEKERRR 761
QY 979 -----RQHRQFVELREAAIFLQRIW-----RELFQKLLAAETARLQRS 1020
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 762 SELOQBEERAHQOQEEBQERDPTWO--WQAEKSRGRGURUSARPLREORERQRAE 819
QY 1021 QKQAAAAYIQWRTYQIGRIQRH-----BFLQRDLIMFVQRRMRKSMLEQ--RKE 1073
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 820 ERQQRQRFLEPEEKEQGRQRREREKELQLEEEQL---QRRERAQQLQEEEDGLQE 876
QY 1074 FQQLKRAINTQQRWAKLSMRKCNADYIALRSSVLKQVAYRKATIQWRIDRNHYISLRK 1133
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 877 DQERRRQQRDQKRWQLEERKRRH-----TLYAKPALQEQ-----LRK 918
QY 1134 NVICLQRLRAIMKQREENVYLNLENASILVQKYMRCQMIQDRNAYLRTRKCIINVQ 1193
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 919 EQQLLOEEREL--QREERKRRQEQ-----EQRYREEQQLQEEBQLLEER---EKR 968
QY 1194 RWRATLQMRERNKYLHLOTTXRI-----QIKFRAKREMKQRAEFLQKLVTVVQK 1248
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 969 RQERERQYRKDK---LQKEEQLGEEPEKERRQREKKYREEBELQOEEBQLLEEE 1024
QY 1249 RRALLQMKEROEYVILHREVITIKLQRFHQAQSMFRAKVRGTQAVSCLQMWENHL 1308
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1025 RE-----KRRQEW--BQVKKDELQOEBQLLEERKRR-----LQERERQY- 1067
QY 1309 LFKRERNFQLOQAAITLQRRYRRLNWKLSYALQKQAATI-----QTRYAKKA 1363
DB |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1068 ---REEBELQOEEBQLGEEETRRQRLERQYRKEEBELQOEBQLLEEBEKEKRRERE 1124
QY 1364 MQ-KQVLYQKQREAIKVVQRYRGNLENMKOI-----EYQKQQAQVIRLQ-----KW- 1411

```

```

DB 1125 RQREEEELQOEEBQLLEEBEKEKRRQRLERQYREEBELQOQKQRYRDSQSDLKWQ 1184
QY 1412 W-----RSIRDMRL-CKA-----GYRRI-----RLSSLSIQ-----RKW 1439
DB 1185 WPEPENAVRDNKVKYCKGRENQFQLEDQSQRVDRQSOQDLQHLGEOQERDRERRW 1244
QY 1440 RATVQARR-----QREIFLSTIRKVRLMQAFIRATLLMRQQRREPEMGRRAVVIR 1491
DB 1245 Q---QANRHFPEEQLEEREQEKAKRDRKQ---EEKQLLEEREE---KRRQET-DR 1294
QY 1492 REFARCAMLKAFODYQLIOSSVILVQRFKFRANSM--KQARQEFVQLRTIAVHLQOKE-- 1547
DB 1295 KFRREELQLEEREQFLRQEQ---RDRKFRREELHQEQGRKFLREERQRLREERKELK 1351
QY 1548 RGRKLMIQRNCFOLLRCRMPGFQARAGFMARKFQALMTPEMMDLRLQKRAAKVIQRY 1607
DB 1352 EQQLRLER--EQLR-----QDRDKF-----REEQQLSRQDRDKFRREE 1392
QY 1608 WRGYLIRRQKHQGLDIRKIAQLRQE 1635
DB 1393 QQ---VRQERERKFLREBEQ---QLRQE 1414

RESULT 14
GOA3_HUMAN
ID GOA3_HUMAN STANDARD; PRT: 1498 AA.
AC Q08378; O43241; Q86XW3; Q8TDA9; Q8WZA3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A member 3 (Golgin-160) (Golgi
complex-associated protein of 170 kDa) (GCP170).
GN GOLGA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), PHOSPHORYLATION, AND SUBCELLULAR
RP LOCATION.
RC TISSUE=Pancreatic carcinoma;
RX MEDLINE=97442456; PubMed=9295333;
RA Misumi Y., Sonda M., Yano A., Fujiwara T., Ikehara Y.;
RT "Molecular characterization of GCP170, a 170-kDa protein associated
with the cytoplasmic face of the Golgi membrane.";
RL J. Biol. Chem. 272:23851-23858 (1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), MUTAGENESIS OF ASP-59; ASP-139 AND
RP ASP-311, SUBCELLULAR LOCATION, PHOSPHORYLATION, AND CLEAVAGE BY
RP CASPASES.
RX MEDLINE=20253169; PubMed=10791974;
RA Mancini M., Machamer C.E., Roy S., Nicholson D.W., Thornberry N.A.,
RA Casciola-Rosen L.A., Rosen A.;
RT "Caspase-2 is localized at the Golgi complex and cleaves golgin-160
during apoptosis.";
RL J. Cell Biol. 149:603-612 (2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Kondo M., Matsukuma S., Hirose F., Matsuda M., Yoshihara M.,
RA Misumi Y., Aida M., Ikehara M., Sutou S.;
RT "Molecular characterization of Mea-2/golgin-160/GCP170 gene encoding a
Golgi membrane associated protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 780-1348 FROM N.A.
RC TISSUE=liver;
RX MEDLINE=93301617; PubMed=8315394;
RA Fritzler M.J., Hamel J.C., Ochs R.L., Chan E.K.L.;
RT "Molecular characterization of two human autoantigens: unique cDNAs
encoding 95- and 160-kD proteins of a putative family in the Golgi
complex.";
RL J. Exp. Med. 178:49-62 (1993).

```

<p>[5] SEQUENCE OF 799-1448 FROM N.A. (ISOFORM 1). TISSUE=Muscle; MEDLINE=22388257; PubMed=12477932; Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). [6] DIMERIZATION, AND SUBCELLULAR LOCATION. MEDLINE=2229380; PubMed=42130652; Hicks S.W., Machamer C.E.; RT "The NH2-terminal domain of Golgin-160 contains both Golgi and nuclear targeting information." J. Biol. Chem. 277:35833-35839(2002). CC -!- FUNCTION: Golgi auto-antigen; probably involved in maintaining Golgi structure. CC -!- SUBUNIT: Homodimer. CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein tightly associated with Golgi stacks. CC -!- ALTERNATIVE PRODUCTS: Event-Alternative splicing; Named isoforms=2; Name=1; IsoId=Q08378-1; Sequence=Displayed; Name=2; IsoId=Q08378-2; Sequence=VSP 007728, VSP_007729; Note=No experimental confirmation available. May be due to an intron retention; CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains. CC -!- PTM: Cleaved into rod-like protein with coiled-coil domains. CC -!- PTM: Phosphorylated. CC -!- SIMILARITY: Belongs to the golgin family. CC -!- CAUTION: Ref.1 sequence differs from that shown due to a frameshift in position 1378. ----- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a>). ----- DR EMBL; D63997; BAA23661.1; ALT FRAME. DR EMBL; AF485338; AAL93149.1; -- DR EMBL; AB027133; BAB71953.1; -- DR EMBL; L06148; AAK35921.1; ALT_INIT. DR EMBL; BC048203; AAK48203.1; -- DR PIR; JH0820; JH0820. DR Genew; HGNC:4426; GOLGA3. DR MIM; 602581; -- DR GO; GO:0017119; C:Golgi transport complex; IDA. DR GO; GO:0005215; F:transporter activity; NAS. DR GO; GO:0006891; P:intracellular transport; NAS. KW Antigen; Golgi stack; Coiled coil; Phosphorylation; KW Alternative splicing. DOMAIN 20 95 PRO-RICH.</p>	<p>FT DOMAIN 172 257 FT DOMAIN 394 1459 FT DOMAIN 538 744 FT DOMAIN 1231 1369 FT SITE 59 60 FT SITE 139 140 FT SITE 311 312 FT VARSPLIC 1382 1390 FT VARSPLIC 1391 1498 FT MUTAGEN 59 59 FT MUTAGEN 139 139 FT MUTAGEN 311 311 FT CONFLICT 159 159 FT CONFLICT 609 609 FT CONFLICT 746 746 FT CONFLICT 785 785 FT CONFLICT 932 932 FT CONFLICT 1017 1017 FT CONFLICT 1281 1281 FT CONFLICT 1315 1315 FT CONFLICT 1443 1443 SQ SEQUENCE 1498 AA; 167353 MW; 4B95CDFD3D64667 CRC64; Query Match 2.9%; Score 274; DB 1; Length 1498; Best Local Similarity 18.7%; Pred. No. 1.8e-07; Matches 297; Conservative 278; Mismatches 531; Indels 482; Gaps 70; QY 263 BEGPSPP-----RPTKSAIHDLK---DIKLVGSPLRKYSESMDL 301 DB 68 QNGTPPPPPPSLDDPTTSPGPDAGFVAGFVHNLKSGQTSAGSVRKEALQSLR-L 126 QY 302 SLLSPQKVAIQGS-MENLNEMKIRSIQNRYYQEQIQIKAKDLNLSSSSEA----- 353 DB 127 SLPQETQLCSTDSPLEKEQVR-LQARKWLEBQLQYEVKEQQRSSQPATKRLFS 185 QY 354 -----SLAQOQEFNHSI-----LAQSRFNLHVGRKSVKSPVN 392 DB 186 TLDPELMNPNLPRASTLMTKEYSFLRTSVPRGPVGLSLFPAHPREKTSKSKTRS 245 QY 393 PHKRSHELSPSDAPN-----ESLYRNET-----VAISPPKQVEDTTLF-R 435 DB 246 LADYRTEDSNAGNGVNPADPTKGLSKQNRSSAASVSEISISPDTRDLNLTSLAGD 305 QY 436 SAAPANASARSSSAHAWPHAQSKFKLAQTMKMPATPRKVRDTSIQPSVKLYDSELY 495 DB 306 SVSEVDGNDSDSSSYSSASTGTGYLLSKTVGTQDTP-----Y 343 QY 496 MQTCINPDPPFAATTIDPFLASTWYLDQAVDRHQADFKNMINALVSTPADLDLNNKI 555 DB 344 M---VNGQEIFADT-----LGQPPSIKDVLQAAAEHQ 373 QY 556 DVGKLFN-EVRNK-----ELVAPTKEQSMNYLTXYRLETLKAAVELFFSBOM 604 DB 374 DQGEVNGEVRSDSDSSVSLESAATQEM-----LQVLK-----EKM 415 QY 605 RLPCSKVAVYVN-KOALRIRSDRNHLHDVWQMORTILELLCFNPLWRLGLEWVFEKIQ 663 DB 416 RLEGQLEALSLEASQALK-----RUMAKVEDMQRM-----EKAE 437 QY 664 MQSNRDIIVGLSTFILNRLFRNKCEQRYSKAYTLTEEVAYETKHH--SLQKILFLPFLD 721 DB 438 LQA--QLAALSTKLQAV---ECSSHQQQRQDLSSE-VDTLKQSCWDLERAMTDLQNNL 491 QY 722 QAKOKRIYVKNPCLFVKPSPKHETKDILLRFSSELLANIGDITRELRRLGYVLOHQTFP 781 DB 492 EAKNASLASNNDLQVABEQYQ-----RUMAKVEDMQRM-----526 QY 782 DEFYAFNNLAVDLRDGVRL--TRVVEVILLDDLTQQLV--PAIRLQRI---FNVKL 834 DB 527 ----LSKNTVHDLRQQMTALQSQVQLERTTLTSLKASQAEISSIQLSVRQWQQQL 582</p>
--	---





```
FT DOMAIN 789 813 IQ 2.
FT DOMAIN 814 836 IQ 3.
FT DOMAIN 837 861 IQ 4.
FT DOMAIN 862 884 IQ 5.
FT DOMAIN 885 913 IQ 6.
FT DOMAIN 914 1237 COILED COIL (POTENTIAL).
FT DOMAIN 1314 1443 COILED COIL (POTENTIAL).
FT DOMAIN 1685 1790 DILUTE.
FT NP BIND 163 170 ATP (POTENTIAL).
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1758 1758 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1853 AA; 215594 MW; 503E93D48CA6B766 CRC64;

Query Match
Best Local Similarity 18.7%; Pred. No. 3.3e-07;
Matches 290; Conservative 250; Mismatches 564; Indels 449; Gaps 70;

QY 518 TWYLDQAVDRHQ-----ADPKWLNALVSIIPADLDADLNKNDVGLKFNVEVRN 566
Db 358 TIFCDLMGVDBEMCHWLCHRKLATATETVTKPSIKLOATNARDALAKHIYAKLFNWIVD 417
QY 567 --KELVAVPKESQSNYLTKXLETILRKAARELFFSEQMRLPCSKVAVVYVVKQALRIS 624
Db 418 HYNQALHSAVKQHSFGLVDIYGFETFEINSFEQF-----CIN--YANE--KLQQ 463
QY 625 DRNIHL-----DVMQRTILELL-----CNPLWRLRLGLEVPFGEKIQMSNRDVG 672
Db 464 QFNHMFVKLEQEBYMEQIPWTLIDFYDQPCINLIESKILGILDLLEDECKMPKGTDDTW 523
QY 673 LSTFILNRLFRNCK-----EQRYS-KAY-----TLTBEYAEATIK 706
Db 524 AQKLYNTHL--NKALFEPRMSNKAIIKHFAKVQEYQCEGFLEKNKQDVFEEQIKVLK 581
QY 707 KHSLOKILFLPLDQAKOKRIVK-----HNPCLFVVKSP-----HKETKDI 748
Db 582 SSKFK-----MLPELFODDEKAISSPTSATSGRTPLTRVPVKTPGPGQTAKHEKHTVGH 637
QY 749 LRFSSSELLANIGDITRELRRLGYVLQHRQTFLDDEPDYAFNNLAVDLRDGVRLTR---VV 805
Db 638 QFNSLHLL-----METLNATPHVYVCIKPNDFKF-PPTFDEKRAVQQLRACGVL 687
QY 806 EVILL--RDLTRQLRVPALSRQIRFNVKLAG-----ALGEANFOLGGDI 850
Db 688 ETIRISARGFPSPRTWYQEPFSRYRVLKQKQVLDGRKQTCNVLKILDKOKYQFG---744
QY 851 AAQDIVDGHREKTLSSLWOLIV--KFRSPKFAAATVLOK---W-WRRHWL-----HVV 898
Db 745 -----KTKIFFAGQAVYLEKLRADKLRAACIRIQKTINGWLLRKRYLCMQRAAIT 795
QY 899 IQRRIR-----HKELMRHRRAATVIQAVPRGHOMERYVKLFKTERTOAAIILQKFTRRY 952
Db 796 VQRYVRGYQARCYAKFLRRTKAATTIQKYMVMYVVRRYKI---RRAATIVIQSYLRGY 851
QY 953 LAQ--KQIYQSYHSIIITQ---RWRAQOLGROHRORFVELREAAIFLQRIWRRRLPAK 1006
Db 852 LTENRYRKILREYKAVI-IQKRVGW---LARTHVKRTM---KAIYVLCQCCFR-----898
QY 1007 KLLAAAEATLARLQRSOKQQAASVYIQWRTYQLGRIOREHFLR--QRDLIMFVQRRMRGK 1064
Db 899 -MMAKEDVKKLEARS-----VERYKKLHGMENKIMQLQKVD--937
QY 1065 WSNLQORKEFOQLKRAAINIQWRRAKLSNRKCNADYLALRSSVLKQVAYRKATIO----1120
Db 938 ----EQNKDYKCLMEKLTNLEGVYNSBETELRNDVERLQLSBEAAKVATCRVLSLQEEIA 993
QY 1121 -MEIDRNHYYSLRKNVICIQRLRAIMKWRBQRENVYLRNANSLVLYOKRYMRQMIQDR 1179
Db 994 KLRKOLEQTRSEKKS1---EERADKYQETDQLVSNLKEEN-TLLQKEKETLNHRIVEQA 1049
QY 1180 NAYLRT--RKCIINVORRRATLQMRERKRNVIHLQTTTKRIQIKF-RAKREMKIQ----1232
Db 1050 KEMTETMERKL---VEETKQLELDLNDERLRYONLNEFSRLEERYDDLKEEMTLMNVP 1106
```

```
QY 1233 -----RAEFLQL-----KKVT-----LVQKRRRALL 1254
Db 1107 KPGHKRTDSTHSSNESEYTFSEFAETEDIAPRTEPIEKVPDMSLFLKQKRVTELE 1166
QY 1255 QMRKEROYVILHREVITIKLQRRPFAQKSMRFRKRGCTQAAVSCLOMWRNHL--LRKR 1312
Db 1167 QEKQLMODBLDRKE-----EQVFRSKAKEBERPQIRGAELYESLK--RQSESENKK 1217
QY 1313 ERNSFLQROAATTLQRR-----YRARLANMIKOLKSYAOLKQAAITFIQRYRA 1360
Db 1218 LKNELNELRKA--LSEKSAPEWAPAGAPAYRVLMEQLTSVSEELDVREEVLI---LRS 1271
QY 1361 KKAMQKQVVLYQKQREAI-----KVORRYRGNLBMWKQIEVYQKQQA 1404
Db 1272 QLVSKQEAIQKDDKNTMTDSTILLEDVQKMDKGBIAQAYIGLKETNRLE-----1323
QY 1405 VIRLOKXWBSIRD-----MRLCKAGYRIRLSSLSIQKWRATVQARRORBEIFLST 1455
Db 1324 -SOLQSKSHENEAELRGEIQSLKEENRQOQLAQNLOLPPEARIASLQHEITRLT 1382
QY 1456 IRKVRMLQAF-----IRATL-----LMRQORREPEMKRRAAVVIQRRFRAR 1496
Db 1383 NEWLIFEELYADDPKYQSYRISLYKGMIDLMEOLEKQDQKTVKLLKQKVPFAKKIGELE 1442
QY 1497 CAMLKARQDYQLIQSSV--ILVQRKFRANRSMKQARQEFVQLRTTIAVHLQOKFRGKRLMI 1554
Db 1443 VQGMENISPGQIIDEPVPIPRKEKDFQGMLEYKREDEQKLVKNLILELPRGVAVNL 1502
QY 1555 EQNRCFQILRCSMPGQARARGFMA-----KRFQALMTPEMMDLIRQKAAKVIQ 1606
Db 1503 -----IPGLFAYILFWCVRHADYLDNDQKVRSLTSTINSI-----KKVLKK 1544
QY 1607 -----YMRGYLIR-----RRQKHGGLDIRKRIAOLRQEAKA 1638
Db 1545 RGDDEPETSFVLSNTRCFHLCLKQYSGEGFMKHNSTRQNEHCLTNF--DLAERYQ---V 1599
QY 1639 VNSVRCKV-QEAVRFLRG--RFIASDAL-----AVLSQDLRLSRTV 1676
Db 1600 LSDLAIQIYQLVRVLENIQLPMIVSGMLEHETIQGVSGVKPTGLRKRTSSIADEGTYL 1659
QY 1677 PHLLMWCSEFMSTFCYGINAQAIRSEVDKOLIERCSRIILN-----LARY 1721
Db 1660 DSILRQLNSFHSVWCOHGMDBELIKQVVKOMFYIVGAITINNLLLRKDKMCSWSKGMQIRY 1719
QY 1722 NSTTV-----NTFOEGG-----LVITIAQML--LRWCDKDSIEFNTICTLI 1759
Db 1720 NVSQLEEWLRDKNLMNSGAKETLEPLIQAAQLLVKKTKTDDDAEAICSMCNAL 1772
```

Search completed: September 13, 2004, 17:41:15  
Job time : 38 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:37:14 ; Search time 103 Seconds  
(without alignments)  
5700.771 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVAPDTLCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.25.\*  
1: sp arChaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	9514	100.0	1861	5	O01401	O01401 drosophila
2	9482	99.7	1861	5	Q9VC45	Q9vc45 drosophila
3	4200	44.1	877	5	Q8X666	Q8x666 drosophila
4	1374	14.4	3122	11	Q8CJ27	Q8cj27 mus musculus
5	1360	14.3	3477	4	Q8IZT6	Q8izt6 homo sapien
6	868	9.1	1142	4	Q8N4D1	Q8n4d1 homo sapien
7	733.5	7.7	825	4	Q8IUL2	Q8iul2 homo sapien
8	733.5	7.7	826	4	Q8IZJ8	Q8izj8 homo sapien
9	688	7.2	882	4	Q8IZJ9	Q8izj9 homo sapien
10	408.5	4.3	726	4	Q9NWS1	Q9nws1 homo sapien
11	370.5	3.9	1110	10	Q84Z28	Q84z28 oryza sativ
12	361.5	3.8	1186	5	O17666	O17666 caenorhabdi
13	324.5	3.4	1088	10	O49705	O49705 arabidopsis
14	314	3.3	1909	5	Q25893	Q25893 plasmodium
15	300	3.2	1374	10	Q9SMY9	Q9sm9 arabidopsis
16	290.5	3.1	285	4	Q9NVT6	Q9nvt6 homo sapien

17	281	3.0	575	11	O88482	O88482 mus musculu
18	272.5	2.9	2473	11	Q9QZ84	Q9qz84 mus musculu
19	271	2.8	1705	5	Q9U0S7	Q9u0s7 mytilus gal
20	270	2.8	2009	11	Q8CJ40	Q8cj40 mus musculu
21	268.5	2.8	2245	5	Q86A36	Q86a36 dictyosteli
22	267.5	2.8	1596	5	Q8IJ44	Q8ij44 plasmodium
23	262.5	2.8	1849	5	Q9U487	Q9u487 loligo peal
24	261.5	2.7	446	5	Q9GYC6	Q9gyce leishmania
25	259.5	2.7	1919	4	Q7Z2L4	Q7z2l4 homo sapien
26	257.5	2.7	1800	5	Q8MKM1	Q8mkmi drosophila
27	254.5	2.7	1837	3	Q74424	Q74424 schizosacch
28	254	2.7	3267	5	Q8IBZ0	Q8ibz0 plasmodium
29	252.5	2.7	1845	11	Q7TQL2	Q7tql2 mus musculu
30	252.5	2.7	2115	4	Q14980	Q14980 homo sapien
31	252	2.6	1764	5	Q9N9U8	Q9n9us leishmania
32	251.5	2.6	1792	5	Q97417	Q97417 drosophila
33	251.5	2.6	1792	5	Q9V4M6	Q9v4m6 drosophila
34	251.5	2.6	2094	11	Q8OY35	Q8oy35 mus musculu
35	250.5	2.6	1416	4	Q9BZF9	Q9bzf9 homo sapien
36	250.5	2.6	1416	4	Q9HCL1	Q9hcl1 homo sapien
37	250.5	2.6	1982	11	Q8OZE6	Q8oze6 mus musculu
38	250.5	2.6	1995	4	Q7Z406	Q7z406 homo sapien
39	250	2.6	2048	5	Q9VCD1	Q9vcd1 drosophila
40	250	2.6	2101	4	Q14981	Q14981 homo sapien
41	248.5	2.6	1302	4	Q7Z2L3	Q7z2l3 homo sapien
42	245.5	2.6	1449	6	Q9BG87	Q9bg87 bos taurus
43	245	2.6	1401	5	Q7YU40	Q7yu40 drosophila
44	242	2.5	3616	13	Q9W6V0	Q9w6v0 gallus gall
45	239.5	2.5	1489	5	Q8T8Q5	Q8t8q5 drosophila

## ALIGNMENTS

RESULT 1  
O01401 ID O01401 PRELIMINARY; PRT; 1861 AA.  
AC O01401;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Microtubule associated protein.  
GN ASP OR CG6875.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON R;  
RA Saunders R.D.C., Avides M.C., Howard T.I.A., Gonzalez C.,  
RA Glover D.M.G.; J. Cell Biol. 0:0-0(0).  
RL EMBL; U95171; AAB51540.1; -.  
DR FIR; T13845; T13845.  
DR FlyBase; FBgn0000140; asp.  
DR GO; GO:0005875; C:microtubule associated complex; IDA.  
DR GO; GO:0008017; F:microtubule binding; IDA.  
DR InterPro; IPR001715; Galponin-like.  
DR InterPro; IPR000046; IQ\_region.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00612; IQ; 21.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00015; IQ; 5.  
DR PROSITE; PS50021; CH; 1.  
DR PROSITE; PS50096; IQ; 5.  
SQ SEQUENCE 1861 AA; 219558 MW; 08A943D5B335EFF45 CRC64;

Query Match 100.0%; Score 9514; DB 5; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELVWSVLEVA CKETIQLIDNRNFRKEVMIILKSKSNQPVKPRKPTVGKTLQKSP 60  
Db 1 MELVWSVLEVA CKETIQLIDNRNFRKEVMIILKSKSNQPVKPRKPTVGKTLQKSP 60  
QY 61 GAGTKMSVVSAAVQCKRMSAAAPSPKOTWRVTAAPSRAAAHAPPOAPLVEKNVYKT 120  
Db 61 GAGTKMSVVSAAVQCKRMSAAAPSPKOTWRVTAAPSRAAAHAPPOAPLVEKNVYKT 120  
QY 121 POEPPVYISQPSRLKENLSPMTGNLLDVINDIRFPLTETRGKQATIPFNDLAAPT 180  
Db 121 POEPPVYISQPSRLKENLSPMTGNLLDVINDIRFPLTETRGKQATIPFNDLAAPT 180  
QY 181 PTLKGNVKSANDMRPRITPDDEDOPATNKFDVKHSETIMISLDTLDCSIDGQPH 240  
Db 181 PTLKGNVKSANDMRPRITPDDEDOPATNKFDVKHSETIMISLDTLDCSIDGQPH 240  
QY 241 PLNKTTTIVHATHTRALACIHEEBGSPPTPTKSAIHLDKRDIKLVGSPLRKYSKMD 300  
Db 241 PLNKTTTIVHATHTRALACIHEEBGSPPTPTKSAIHLDKRDIKLVGSPLRKYSKMD 300  
QY 301 LSLILSPOTKVAIOGSMPLNEMKIRTEONRYVBOOOIOIKAKDLNSSSSSEASLAGOE 360  
Db 301 LSLILSPOTKVAIOGSMPLNEMKIRTEONRYVBOOOIOIKAKDLNSSSSSEASLAGOE 360  
QY 361 FLFNHSEILAQSSRFNLHEVGRKSVGSPVKNPKRSHLSLSPDAPSNESLYRNETVAI 420  
Db 361 FLFNHSEILAQSSRFNLHEVGRKSVGSPVKNPKRSHLSLSPDAPSNESLYRNETVAI 420  
QY 421 SPPKKQVBDTTLPRSAAPANASARSSSAHAWPHAQSKPKLAQTMSLMCKPATPRKVRD 480  
Db 421 SPPKKQVBDTTLPRSAAPANASARSSSAHAWPHAQSKPKLAQTMSLMCKPATPRKVRD 480  
QY 481 TSIOPSVKLYDSSELYMOTCINDEPFAATTIDIFLASTMYLDEQAVDRHQADPKKMINAL 540  
Db 481 TSIOPSVKLYDSSELYMOTCINDEPFAATTIDIFLASTMYLDEQAVDRHQADPKKMINAL 540  
QY 541 VSPADLDADLNKIDYKGLFNEVRNKELVVAFTKEEQSMNYLTKYRLETIRKAAVELEF 600  
Db 541 VSPADLDADLNKIDYKGLFNEVRNKELVVAFTKEEQSMNYLTKYRLETIRKAAVELEF 600  
QY 601 SEQNLPCSKVAVYVKNQALRISDRNHLVDVWQRTILELLICFPLMLRGLGVFGE 660  
Db 601 SEQNLPCSKVAVYVKNQALRISDRNHLVDVWQRTILELLICFPLMLRGLGVFGE 660  
QY 661 KIOMOSNRDVLGLSTFILNRLFRNKCEQYSAYTLTEEYAEYTIKKHSLOKILFLLPFL 720  
Db 661 KIOMOSNRDVLGLSTFILNRLFRNKCEQYSAYTLTEEYAEYTIKKHSLOKILFLLPFL 720  
QY 721 DOAKOKRIVKHNPCLFVKKSPHKETKDILRPSSELLANTGDTREIRLRYVLOHROTF 780  
Db 721 DOAKOKRIVKHNPCLFVKKSPHKETKDILRPSSELLANTGDTREIRLRYVLOHROTF 780  
QY 781 LDEFYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPALSRQLRIPNVKLALCALG 840  
Db 781 LDEFYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPALSRQLRIPNVKLALCALG 840  
QY 841 EAMFOLGDDIAODIVDGHREKTLISLLWQIYKFRSPKFAAATVLOKWRHRLHVLQ 900  
Db 841 EAMFOLGDDIAODIVDGHREKTLISLLWQIYKFRSPKFAAATVLOKWRHRLHVLQ 900  
QY 901 RRIHKKELMRHRAATVIOAVFGRHOMRKVKVFLKFTERTQAAIIOKFTREYLAOKOLYQ 960  
Db 901 RRIHKKELMRHRAATVIOAVFGRHOMRKVKVFLKFTERTQAAIIOKFTREYLAOKOLYQ 960  
QY 961 SYHSIITQWRRAQQLGROHROFVELREAAIFLQRIWRRRLFAKLLAAAEATLQRS 1020  
Db 961 SYHSIITQWRRAQQLGROHROFVELREAAIFLQRIWRRRLFAKLLAAAEATLQRS 1020  
QY 1021 OKQOAAASYLOMOWETTYOLGRIORHBEFLORDILIMFVORMSKWSMLBORKEFOOLKA 1080  
Db 1021 OKQOAAASYLOMOWETTYOLGRIORHBEFLORDILIMFVORMSKWSMLBORKEFOOLKA 1080  
QY 1081 AINTQORWRAKLSMRKCNADYLALRSVLQVAYRKATIOMRIDRHHYSLRKNVCLQ 1140

Db 1081 AINTQORWRAKLSMRKCNADYLALRSVLQVAYRKATIOMRIDRHHYSLRKNVCLQ 1140  
QY 1141 RLRATMKRREORENVILRLNASILVQKYRMRQOMIQRNAYLRTKCIINVQRWRATL 1200  
Db 1141 RLRATMKRREORENVILRLNASILVQKYRMRQOMIQRNAYLRTKCIINVQRWRATL 1200  
QY 1201 QMRERKNYLHIQTTTKGRIQKFRAKREMKQRAEFLQKKVTLVVOKRRALLQMRKER 1260  
Db 1201 QMRERKNYLHIQTTTKGRIQKFRAKREMKQRAEFLQKKVTLVVOKRRALLQMRKER 1260  
QY 1261 QBYLHLREVTIKLORRPHAQSMRPMRAKYRGTOAAVSCLOWHNRHLLRKRNSFLQ 1320  
Db 1261 QBYLHLREVTIKLORRPHAQSMRPMRAKYRGTOAAVSCLOWHNRHLLRKRNSFLQ 1320  
QY 1321 RQAATLQRRYARLANMIKOLKSYAQLKQAAITITQTRYAKKAMOKOVLYQKQREAIK 1380  
Db 1321 RQAATLQRRYARLANMIKOLKSYAQLKQAAITITQTRYAKKAMOKOVLYQKQREAIK 1380  
QY 1381 VORRYRGNLEWRKOIEVYQKQOAVIRLOKWRSTIRDMELCKAGYRIRLSLSIORKWR 1440  
Db 1381 VORRYRGNLEWRKOIEVYQKQOAVIRLOKWRSTIRDMELCKAGYRIRLSLSIORKWR 1440  
QY 1441 ATVOARRORREIFLSTIRKVRMLQAFIRATILMRQORREFEMKRAAAVVIQRRFRACML 1500  
Db 1441 ATVOARRORREIFLSTIRKVRMLQAFIRATILMRQORREFEMKRAAAVVIQRRFRACML 1500  
QY 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLOKFRGKMLTEORNC 1560  
Db 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLOKFRGKMLTEORNC 1560  
QY 1561 QLLRCSMPGFOARAGFMARKFOALMTPEWMDLIROKRAAKVIQRYWGYLIRRRQKHQ 1620  
Db 1561 QLLRCSMPGFOARAGFMARKFOALMTPEWMDLIROKRAAKVIQRYWGYLIRRRQKHQ 1620  
QY 1621 GLLDTRKRIAOLROBAKAVNSVRCKVQBAVFLRGRFTIASDALAVLSOLDRLSRTVPHLL 1680  
Db 1621 GLLDTRKRIAOLROBAKAVNSVRCKVQBAVFLRGRFTIASDALAVLSOLDRLSRTVPHLL 1680  
QY 1681 MWCSEFMSTFCYGINMAQAIRSEVDKOLIERCSRILNLARVNSTVTMTFOGGVTTAQM 1740  
Db 1681 MWCSEFMSTFCYGINMAQAIRSEVDKOLIERCSRILNLARVNSTVTMTFOGGVTTAQM 1740  
QY 1741 LLRWCDKDSIEFNTLCTLIWVFAHCPKPKRIIHDYMTNPEAIYVRETCKLVARKEKMKQ 1800  
Db 1741 LLRWCDKDSIEFNTLCTLIWVFAHCPKPKRIIHDYMTNPEAIYVRETCKLVARKEKMKQ 1800  
QY 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPDGFIIRSPYTFISSVAFDTILCKLOIDM 1860  
Db 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPDGFIIRSPYTFISSVAFDTILCKLOIDM 1860  
QY 1861 F 1861  
Db 1861 F 1861  
RESULT 2  
QYVC45 PRELIMINARY; PRT; 1861 AA.  
AC QYVC45; PRELIMINARY; PRT; 1861 AA.  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE CG6875-PA.  
GN ASP OR CG6875.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wauson D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
[2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Parasag V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield B.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Muncall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;

Db 781 LDEFYAFNNLAVLDLGRVLRTRVMEVILLRDDLTQLRVPAISRLQRIENVKLALGALG 840

Qy 841 EANFOLGGDIAACDIDVGHREKTLISLLWOLIVKPSPEKFAAAATVLOKWRHSHLHVLIQ 900

Db 841 EANFOLGGDIAACDIDVGHREKTLISLLWOLIVKPSPEKFAAAATVLOKWRHSHLHVLIQ 900

Qy 901 RRIIRKELMRHRAATVIAVFRGHQMRKYVKLFKTERTOAAIILQKFTRRYLAQKQLYQ 960

Db 901 RRIIRKELMRHRAATVIAVFRGHQMRKYVKLFKTERTOAAIILQKFTRRYLAQKQLYQ 960

Qy 961 SYHSIIITTORWRAOQLGHORFVELREAAIFLQRIWRRLFAKLLAAAEARLORS 1020

Db 961 SYHSIIITTORWRAOQLGHORFVELREAAIFLQRIWRRLFAKLLAAAEARLORS 1020

Qy 1021 QKQAAAASYIQMWRTYQLGRTORHEFLFORDLIMFVQRMESKWSLBOKEFOOLKKA 1080

Db 1021 QKQAAAASYIQMWRTYQLGRTORHEFLFORDLIMFVQRMESKWSLBOKEFOOLKKA 1080

Qy 1081 AINIQRWRKLSMRKCNADYLALRSSVLKVOAYRKATIQMRIDRNHYYSLEKNVICLOQ 1140

Db 1081 AINIQRWRKLSMRKCNADYLALRSSVLKVOAYRKATIQMRIDRNHYYSLEKNVICLOQ 1140

Qy 1141 ELRAITMKREOENYILRNASILVQKRYRMEQMIODENAYLTKCIINVQREWRATL 1200

Db 1141 ELRAITMKREOENYILRNASILVQKRYRMEQMIODENAYLTKCIINVQREWRATL 1200

Qy 1201 QMRREKNYLHLQTTTKRIQIKFRAKREMKQABFLQKKVTLVVQKRRALLQMKER 1260

Db 1201 QMRREKNYLHLQTTTKRIQIKFRAKREMKQABFLQKKVTLVVQKRRALLQMKER 1260

Qy 1261 QBYLHLREVITIKQRFFHAQSMRPMRAKYRGTOAAVSCLOQWHNHLRKERNSTFQL 1320

Db 1261 QBYLHLREVITIKQRFFHAQSMRPMRAKYRGTOAAVSCLOQWHNHLRKERNSTFQL 1320

Qy 1321 FOAATLQRRYARLNMILKLSYALQKQALITQITRYRAKAMQKVLYQKQREALIK 1380

Db 1321 FOAATLQRRYARLNMILKLSYALQKQALITQITRYRAKAMQKVLYQKQREALIK 1380

Qy 1381 VQRRYRGNLEMRKQLEVVQKQOAVIRLQKWRISIRDMRLCKAGYRIRLSLSLQKWR 1440

Db 1381 VQRRYRGNLEMRKQLEVVQKQOAVIRLQKWRISIRDMRLCKAGYRIRLSLSLQKWR 1440

Qy 1441 ATVOARRQREIFLSTIRKVLMOAFIRATLMLRQORREFEMKRAAVVIQRRFRACAML 1500

Db 1441 ATVOARRQREIFLSTIRKVLMOAFIRATLMLRQORREFEMKRAAVVIQRRFRACAML 1500

Qy 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQOKFRGKRLMTEORNC 1560

Db 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQOKFRGKRLMTEORNC 1560

Qy 1561 QLLRCSMPFGQARAGFMARKRFQALMTPENMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ 1620

Db 1561 QLLRCSMPFGQARAGFMARKRFQALMTPENMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ 1620

Qy 1621 GLDIRKRIAGLROBAKAVNSVRCVKQBAVFLRGRFTIASDALVLSOLDLSLSTVPHL 1680

Db 1621 GLDIRKRIAGLROBAKAVNSVRCVKQBAVFLRGRFTIASDALVLSOLDLSLSTVPHL 1680

Qy 1681 MWCSEFMSTFCYGIMAOAIRSEVDKOLLERSIILNLARVNSTVNTVFOGGVLVTAQM 1740

Db 1681 MWCSEFMSTFCYGIMAOAIRSEVDKOLLERSIILNLARVNSTVNTVFOGGVLVTAQM 1740

Qy 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKRRKIIDHYMTNPEAIYVVRTEPKLVARKKMKQ 1800

Db 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKRRKIIDHYMTNPEAIYVVRTEPKLVARKKMKQ 1800

Qy 1801 NARPPPMSTSGYKSOKINFPTCSIPSLPDPFGIIRSPYTFISSVAFDTILCKLQIDM 1860

Db 1801 NARPPPMSTSGYKSOKINFPTCSIPSLPDPFGIIRSPYTFISSVAFDTILCKLQIDM 1860

Qy 1861 F 1861

Db 1861 F 1861

RESULT 3

Q8SX66 PRELIMINARY; PRT; 877 AA.

ID Q8SX66;

AC Q8SX66;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE LD39479p.

GN ASP OR CG6875.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkley;

RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,

RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

RA Miranda A., Mungall C.J., Nunoo J., Pacleeb J., Paragas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Celniker S.;

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL, AY094825; AAM11178.1; --.

DR FlyBase; FBgn000140; asp.

DR GO; GO:0005875; C:microtubule associated complex; IDA.

DR GO; GO:0008017; F:microtubule binding; IDA.

DR InterPro; IPR001715; Calponin-like.

DR InterPro; IPR000048; IQ\_region.

DR Pfam; PF00307; CH; 1.

DR Pfam; PF00612; IQ; 16.

DR SMART; SM00033; CH; 1.

DR PROSITE; PS00021; CH; 1.

DR PROSITE; PS00096; IQ; 3.

SQ SEQUENCE 877 AA; 106401 MW; 328AC8FD21C027CE CRC64;

Query Match 44.1%; Score 4200; DB 5; Length 877;

Best Local Similarity 99.4%; Pred. No. 1.9e-265;

Matches 829; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 634 MORTLELLLCFNPWLRLGLEVVFGKIQMCSNRDIVGLSTFTILNRLFRNKCEQRYSK 693

Db 1 MORTLELLLCFNPWLRLGLEVVFGKIQMCSNRDIVGLSTFTILNRLFRNKCEQRYSK 60

Qy 694 AYLTTEEYAEITKKHSLQKILFLPFLDQAKQKRVKHNPCLFVKKSPHKETKDILLRFS 753

Db 61 AYLTTEEYAEITKKHSLQKILFLPFLDQAKQKRVKHNPCLFVKKSPHKETKDILLRFS 120

Qy 754 SELLANIGDITRELRRLGYVLOHRTFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDD 813

Db 121 SELLANIGDITRELRRLGYVLOHRTFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDD 180

Qy 814 LTRQLRVPALSRLOKIFNVKLALGALGEANFOLGDDIAAODIVDGHREKTLISLLWOLLYK 873

Db 181 LTRQLRVPALSRLOKIFNVKLALGALGEANFOLGDDIAAODIVDGHREKTLISLLWOLLYK 240

Qy 874 FRSPKFAAAATVLOKWRHSHLHVVIQRRIRHKELMRHRAATVIAVFRGHQMRKYVKL 933

Db 241 FRSPKFAAAATVLOKWRHSHLHVVIQRRIRHKELMRHRAATVIAVFRGHQMRKYVKL 300

Qy 934 FKTERTOAAIILQKFTRRYLAQKOLYOSYHSIITIQSWRAQOLGROHROFVELREAAI 993

Db 301 FKTERTOAAIILQKFTRRYLAQKOLYOSYHSIITIQSWRAQOLGROHROFVELREAAI 360

Qy 994 FLQRIWRRRLFAKLLAAAEARLQRSQKQAAASYIQMWRTYQLGRIOHREFTLRQDL 1053

Db 361 FLQRIWRRRLFAKLLAAAEARLQRSQKQAAASYIQMWRTYQLGRIOHREFTLRQDL 420

Qy 1054 IMFVQRMESKWSLBOKEFOOLKRAAINIQSWRAKLSMRKNADYILARSSVLKVOA 1113

```

Db 421 IMPVORMRKSWMLBQKEFOQKRAAINIQORWRAKLSMRKCNADYLALRSSVLKVOA 480
QY 1114 YRKATTOMRIDRHHYSLRKNVICLOQLRAINKMREORENYILRLNASILVQKRYMRQ 1173
Db 481 YRKATTOMRIDRHHYSLRKNVICLOQLRAINKMREORENYILRLNASILVQKRYMRQ 540
QY 1174 QMIQDNAYLRTKCIINQVRRWRATLQMRERKNYLHLQTTTKRIQIKPRAKREMKKOR 1233
Db 541 QMIQDNAYLRTKCIINQVRRWRATLQMRERKNYLHLQTTTKRIQIKPRAKREMKKOR 600
QY 1234 ABFLQJLKVTVLVQKERRALLQMRKQERYLHLREVTIKLORFPHAQKSMRFRKRYGT 1293
Db 601 ABFLQJLKVTVLVQKERRALLQMRKQERYLHLREVTIKLORFPHAQKSMRFRKRYGT 660
QY 1294 QAAVSCLOHWRNHLRKERNSPLOLRQAATLORRYARLNMILKOLKSYAOLKQAAT 1353
Db 661 QAAVSCLOHWRNHLRKERNSPLOLRQAATLORRYARLNMILKOLKSYAOLKQAAT 720
QY 1354 IQTRYAKKAMQOVVLYQKORAILIKVQRRYRGNLEMRKOIEVYQKORQAVIRLOKQWR 1413
Db 721 IQTRYAKKAMQOVVLYQKORAILIKVQRRYRGNLEMRKOIEVYQKORQAVIRLOKQWR 780
QY 1414 SIRDMLCKAGYRRIRLSLSIQKWRATVQARRQREIFLSTIRKVKLMOAFIR 1467
Db 781 SIRDMLCKAGYRRIRLSLSIQKWRATVQARRQREIFLSTIRKVKLMOAFIR 834

RESULT 4
Q8CU27 PRELIMINARY; PRT; 3122 AA.
AC Q8CU27
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Abnormal spindle.
DE CALMBP1 OR ASP.
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster;
RC MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Askham J.M., Springell K., Mahadevan N., Crow Y.J., Markham A.F.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR RMEL; AF533752; AAN46088.1; -.
DR MGD; MGI:1334448; Calmbp1.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00612; IQ; 61.
DR SMART; SM00033; CH; 2.
DR SMART; SM00015; IQ; 55.
DR PROSITE; PSS0021; CH; 2.
DR PROSITE; PSS0096; IQ; 12.
SQ SEQUENCE 3122 AA; 364117 MW; 80DED62DE0F4E7B2 CRC64;

Query Match 14.4%; Score 1374; DB 11; Length 3122;
Best Local Similarity 22.8%; Pred. No. 3.7e-80;
Matches 553; Conservative 369; Mismatches 743; Indels 756; Gaps 76;

QY 1 MELVWSPVLEVACKETLQILDNRNFRKEVMILKSKNQPKVPKRPFTVGKTLQKSP 60
Db 98 ISVTWTPLRGGVREIVTFLVN-DFLKHQAILL-GNAEPPKKKSLWNTSK-----KIPA 151
QY 61 GAGKTMKSV-----VGAQVQKTR-----MSAAAAPPSPQTVRWVTHAPS-RPA 101
Db 152 SSKHTKTSKQNHFNESFTISQDRIIRSPLOPCENLAMEGSSPTEN---KVPTPSIPI 208

```

```

QY 102 AWAHPPQAPLV--EKNVYKTPQEEPVYISOPRSILK-ENLSPMTFCNLLDVIDNLRP-T 157
Db 209 RECQSETCLPFLRESTAYSSLHE-----SENTQNLKQVDASISQTFDFNEEVANETFIN 263
QY 158 PLT---ETGKGQATIPDNLAAMPPTLKGNVKSCAND--MRPRITPDDEDDLED-QPATN 211
Db 264 PISVCHQSEGRKRLTAPN--CSSPLNSTQTHFLSPDSFVNNRYTSDNDLSKMNKVL 321
QY 212 KPFVKGHSITINISLTDLD--CSRIDGQHPHLNKTITIVHATHRALACIHEEGSPSP 269
Db 322 DTRKDPABSVLCESQSTVHEVCQTI--LSPDSFLNDYGLKKGLNFKSV-----NPV 371
QY 270 RPTPKSAIHDLRDILKLVGSPLEKYSKMLSL-----LSPQTKYAIQSGMPLNEMKI 324
Db 372 LSPTQF-----VKDSMGHVQQTGKNEASQDMRINEGLAYTPECQHA---QTPSSRSEKQ 424
QY 325 RSIE-----QNRYYQEQQIQIKAKDNLNSSSSSEASLACQOEFLPNHS 366
Db 425 NPVEVKPHKYDFTKQPKTCFQDAFCHQSKQPKRPFILSATVTTRKPTNAREKL---P 481
QY 367 EILAQSSRFNLH-----EYGRKSVKG---SPVKNP-----H 394
Db 482 EINKDPACKCLEGLVQGRKEVGLSREKGFHPSLPVVEPGVSKALSYRDEVTPATVVVAR 541
QY 395 KRSHSLSFSDAFSNESLYRNETVAISPCKK---QREVD--TLPRSAAPANASARSSA 449
Db 542 KRKSHG-TVGDA--NGKVAEEWMDMCEVKRIHFSPLESTPSTVARTTKKEGLSKRISS 598
QY 450 HAWPHAQSKFKLAQWISLMKKPATPRKVRTDSIQPSVKLYDSELYMQTCINPDPAATT 509
Db 599 -----LERSGLKKMDSSILKTLPSKTKKRRSI---VAVAQSHL---TFIKPLKAAIPR 647
QY 510 TIDPFLASTWYLDQAVDRHQADEFKKWLNALVSI PADLDADLN--NKIDVGKLFNEVRNKE 568
Db 648 HMPFAAKMIFYDERWKEKQEQGFTWNLNLT--PDDFTVKTNVSKVNAASVLGASQH 706
QY 569 LV---VAPTKEEQSMN--YLTKYRLETLRKAAVELFFSEQMRLPCKSVAVYVNVKQALIRS 624
Db 707 KISVPKAPTKEEVSRLAYTASCRLNRLRTACSLFTSEKMKVAKKKVEIEVGRLLVRK 766
QY 625 DRNLHLDVVMQRTILELLCFNPLNLRGLLEVVFGEKIQMOSNRDIVGLSTPILNRLFRN 684
Db 767 DRHLWKDIGQRKVLNWLSSYNPLRIGLETYVFGELIPLADNSDVTGLAMFLNRLWN 826
QY 685 K--CEQRQSKAYTLTEE-YAETIKKHSLOKILFLPFLDQAKQKRVKHNKPCLFVKKSP 741
Db 827 PDIAAEVRHPTVPLPRDGHAAALSKFTLKKLLILCFLDHAKISRLIDHDPCLFCKDAE 886
QY 742 HKETKDILLFSSSELLANIGDITRELRRLGYVLQHRQTFLDEFDYAFNNLAVDLRQVRL 801
Db 887 FKASKELLALAFSRDFLSGEGDLSRHLSTFLGLPVSHVQTPLDDEFDAVTNLAVDLQCGVRL 946
QY 802 TRVVEVILLRDDLTRQLRVPALSRQLRIFNVKLAGALGANFQL-----GGDIAAQDIVD 857
Db 947 VRTVELLTQNNWLSDKLRIPAIRSVQKMHNVDLVLQVLSRGVPLTDEHGAISSKDDVD 1006
QY 858 GHREKTLILLWQLIYKFR-----
Db 1007 RHREKTLGLLWKTALAFQVDISLNDQLKEIDFLKHTHSIKRAMSALTCPESQAITNKQR 1066
QY 876 -----
Db 1067 DKRISGNFRYGDVSQVLLMDVWNAVCAFYNNKVENFTVSFSDGRVLCYLIIHYHPCYVPF 1126
QY 876 -----SKFPH 880
Db 1127 DAICORTSQSACQATGCVSVVNSSSEGGCLDLSLEALDHESTPEMYKELLENEKNF 1186
QY 881 -----AAATVLOKQWR 891
Db 1187 LVNSAARDLGGIPAMIIHSDMSNTIPDEKVVIIYVSLFCARLLDLRKEIRARLIQTWR 1246
QY 892 RHMLHVVIQRRIRH-----KELMRHRAATVIOAVFRGHQMRKY 930

```

```
Db 1247 KYKL-----KDLKHQERDKAARVIQSVVNLFLSRRLQKNVSAALVIQCKWRRVSAQK 1302
Qy 931 VKLEKTER-----TQAAIILQKFTRYLAQKQLYSYH-----963
Db 1303 LKMLKNEKLAKONKSAIOWYRYSTRKRFLKHYSTVLOSIRMKIALTSYKRYL 1362
Qy 964 -SIITIORWRAQQGROHROFVELREAAIFLQ---RIWRRFLFAKLALAAETARLOR 1019
Db 1363 WATVIOQRHWAYLSGRDQIQIFKLKSSSLVIQFWFRWKRKLQTKAAVTLQAFR 1422
Qy 1020 S---OKQ-----QAAASY-----1029
Db 1423 EWHLEKQIRERSAVVIQSWYRMHRELQKVIYIRSCVIVIORVRFCQAKLYKRRKDAL 1482
Qy 1030 -IQOWWRYTQGLRIQRHEFLRDILMFVOREMR-----SKWSMLE 1069
Db 1483 TLQHYRARQKGLAHADYLOKRAATIRLOAFRGMKARHSYRLQIGAACVLQSWRRWQ 1542
Qy 1070 QR-----KFPQOL-----KRAAINIQORWRAKLSMRKCNADYLALRS 1106
Db 1543 ERVFLNLKMKWIKLOAHIRKYQOLQKYKIKKAAITITQFRASISARRVLASYQKTRS 1602
Qy 1107 SVLKVQ-----AYRKATIOMRIDENHY--YSLRK---NVICLOORLAINMKRQ 1151
Db 1603 SVIVLOSACRGWQARKAFRHALASVIKIQSYRYAYICRKTQNFKNATIKLSIVMKQ 1662
Qy 1152 RENYLRLNASILQKRYMRQOMIQDRNAVILRTKCIINVORWRATL-----1200
Db 1663 RKQVLOIRAAALFIORWTRSKLASQKKEYIQVRESCKIQSHFRGCLVRKQLKQCKA 1722
Qy 1201 -----OMRRERNYLHLQTTTKRIQIKPRAKEMKKQRAEFQLKKVTLVVO-----1247
Db 1723 AISLOSYPFMRRTARQYKMKCAALVIOSFYCAVEAQISQRKNFQVQKRAALICLQAAVRG 1782
Qy 1248 -KRRALLQ-----RKEROYLHLREVTIKLORRFAQKSMFMRKAYRGT 1293
Db 1783 CKVRQIQKQSTAAVTIORVFRGHSQRMKYQTMLOSAVKIORWYRAQKVAYDMRTOFLKT 1842
Qy 1294 QAAVSCILQMHWN-----HLLR-----1310
Db 1843 REAVVCLQSAVRCWQVROOLRRQHEAAVKIOSTFEMAVAQOYKLLRAAAVIOQHVRAR 1902
Qy 1311 ---KRENSFLQROAAITLQRRVBARL-----NMIKOLKSYAOLK 1348
Db 1903 AAGKRQHLAYTLQREHAALVFOAAWKGKMLRQIARQHQCAALIOSYYRMHIORRKSIMK 1962
Qy 1349 QAAITIOTRYAKKAMQOVLYQKOREALIKVQRYRGNLEMRKQI-----1395
Db 1963 TAALQIQLCYRAYKVGKEORHLTKAAAVTVLOSAYRG-MKVRKRVAECHKAAVTTQSK 2021
Qy 1396 -----EYQKQOQAVIRLOKWRHSIR-----1416
Db 2022 FRAYRTQKTYTYSALVIOQWYENIKITTOHQEYLNLRRAAVOQAAVYRGITVRRI 2081
Qy 1417 -----DMRLCKGYRRIRLSSUSIQKWRATVOARQOREIFLSTIRKVRML 1462
Db 2082 QHMHMAATLIEAMFKMRSRYRLKMRATAALIIQVRVAYVYLGKIQHEKYRLTLKAKITL 2141
Qy 1463 QAFIR-----ATLLM-----ROOREFEEMKREAAVVIQRRFPCAMLKAR 1503
Db 2142 QAGVGRARVTRVKMHFAATLIIQSHFRGHQQTTFHRLKAAAMVOQRYFAVKEGSAEF 2201
Qy 1504 QDYQLIQSSVILQKFRANRSMQO-----ARQEFVQLRTIAVHLQ 1544
Db 2202 QRYSELRESVLLIIQAAFRGLTRRHLKAMHLAATLIIQRRFTFAMRKFLSKRTAIWQ 2261
Qy 1545 QKFRGKRLMISQRCNFQLLRCMSPGFOARAGFMARXRQAL-----MTPEMM 1592
Db 2262 RQYRA-RLYAKYSRQOLLLEKAVTKIOSSVYGRVVRVQKGRHAAATVIOATFRMHGAYM 2320
Qy 1593 DLIRQKAAKVIQRYWRGYLIRRKQHG-LLDIRKRIALQRLQEKAVNSVRCKVQEAVR 1651
```

```
RESULT 5
Q81ZT6
ID Q81ZT6 PRELIMINARY; PRT; 3477 AA.
AC Q81ZT6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Abnormal spindles.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Ashkam J.M., Springall K., Mahadevan M., Crow Y.J., Markham A.P.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AF509326; AAN40011.1; -.
DR Genew; HGNC:19048; ASPM.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00612; IQ; 71.
DR SMART; SM00033; CH; 2.
DR SMART; SM00015; IQ; 63.
DR PROSITE; PS50021; CH; 2.
DR PROSITE; PS50096; IQ; 10.
SQ SEQUENCE 3477 AA; 409796 MW; 49AE76DC75124980 CRC64;
```

```
Query Match 14.3%; Score 1360; DB 4; Length 3477;
Best Local Similarity 22.7%; Pred. No. 3.6e-79;
Matches 480; Conservative 360; Mismatches 630; Indels 640; Gaps 68;
```

```
Qy 5 WSPVLEVACKETLQIDNFRKVMILKSKSNQPVKNPKRFPPTVGKTLQKSPGTGAGK 64
Db 110 WTPLEGRVREIMTFLVN-DVLKQAILGNABEQKKKRSIWDTI-----154
Qy 65 TMKSVSAAVQOKKMS-----AAAPPKQTMVTAAPSRAAWAHPPQAPLYEKVYK 119
Db 155 -KKKKISASTSHNRVSNIQNVNKTFSQVKDVRSLPQACENLAMEGFPPTENSL- 212
Qy 120 TPQEPYVIS-----QPSRLK-----ENLSPMTPGNLLDV-----IDNLRFT 157
Db 213 ILEENKIPISLSPAFNECHGATCLPLSVRRSTYSLSHASENRELLNVHSANVSKVSFN 272
Qy 158 --PLTET-----RGKQATIPFDNLAAMPPTPKIGNVKSCAN--DMRPRRITPDD 203
Db 273 EKAVTETSFNVSNNVGQGENSKL-----SLTPNCSSTLNTOSQIHFLSPDS 320
Qy 204 LEDQPATNKTFDVKHSETINISLDTLDCSRID---GQHPPLNKTITIVHATHTRALACI 260
Db 321 P-----VNSHGANNLELVTCISDMPMKNSQPVHLESTIAHIIYQKIL---366
Qy 261 HEERGSPPTPTKSAIHD---LKRDIKLYG-SPLRKYSESMDLSLLSPQTKYAIQSGM 316
Db 367 -----SP-----DSFIKDNVGLNQLSESVNFIISPNOFLKDNMAYMCTSQQTKVPL 415
Qy 317 PNLNEMKIRSIQRYQBOQIQIKAKDLNSSSSSEASLACQOQFHLNHSILAQSS-RF 375
```



416	Db	SNENSVQVQSPED---WRKSEVSPRIPEOCQSKSPKAI	FEELVEMKSNYYSFYFKQNNPKF	472
376	Qy	N-LHEVGRKVGSPVKNP-----HRRS-----	HELSEF	403
473	Db	SAVODISSHSHNQPKERPILSATVTKKATCTFENGOTE	INKPKAKELNSAVEHEKVI	532
404	Qy	SDAPSNSL-----YNE-----TVAISPPKKQ---	RVEDTTLPRSAAP	439
533	Db	NNQKEKDFHSYLPIDPILLSKSKYKNEVTPSGSTTAS	VARKRGKSDGSMEDANV-RVAIT	591
440	Qy	ANASARS-SSAHAWPHAQ-----SKKEFLAQWLSMEKP	-----APRKV	478
592	Db	EHTEVRIKRIHFSPPSKTSVAVKTKNVTPLSKRISN	EKLNNKKTKDLSIFRPIISK	651
479	Qy	RDTSIQPSVKLYDSELYMQTCINPDPAATTID-----	PFLASTMYLDEQAVDRHQADF	534
652	Db	TNKETKIIAQASSL--TFIKP-----LKTDIRHPMP	FAAKNMFYDERWKEKQEGFT	704
535	Qy	KWLNALVSI PADLDADLN--NKIDVGKLFNEVRNKL	VV-----APTKEQSNV--YLTKYRLE	589
705	Db	WWLNFILT-PDDFTVKTNISEVNAATLLIGIENGHKIS	VSVPRAPTKEEMSLRAYTARCLN	763
590	Qy	TLRKAAVELFSEOMRLPCSKVAVVNVKQALRTS	DENLHLDTVMQRTILELLCENPLW	649
764	Db	RLRRAACRLFTSEKQWKAIKLEIEISBARRLIVK	RDRHLWKDVGROKVLNLLSYNPLW	823
650	Qy	LRGLEVVFGKIQMOSNRDIVGLSTFILNRLFN--	-KCEQRYSKAYTLTBEYABTIK	706
824	Db	LRIGLETTYGELISLEDNSDVTGLAMFILNELL	WNPDIAAEYRHPTVPHLYRQHEALS	883
707	Qy	KHSLOKILFLPFDQAKQKRIVKHNPCLFVK	SPHKETKDILLRFSSELLANIGDITRE	766
884	Db	KFTLTKLLLVCFDLYAKIRLIDHDPCLCFOA	BFKASKKEILLAFSRDFLSGEGDLSRH	943
767	Qy	LRLGYVLQHQTFELDFEFAFNLAVDLRDGV	ELTVGVILLRDDLTQLQVPAISRL	826
944	Db	LGLGLPNVHQTFDFEFDFAVNLAVDLQCG	VELVMTBELLTQNWDLSKKURIPASRL	1003
827	Qy	QRIENVKLALGALCANFQL-----GGDIAAQDI	VDGHEKRTLSLLWQLIYKFR-----	875
1004	Db	QKMNVDIVQLVKSRTGELSDEHNTILSKD	IVDRHEKRTLRLWKIAFAQVDSLNL	1063
876	Qy	-----	-----	875
1064	Db	DQLKEBIAFLKHTKSINKKTI-SILASHSD	DLINKKKGRKSGSFEQYSENIKLLMDWNAV	1123
876	Qy	-----	-----	875
1124	Db	CAFYNKKNVNTVSFSGRVLVYLIHHYPC	YVDFDAICQRTQTVECTQTSVVLN	1183
876	Qy	-----	-----	880
1184	Db	ESDDSSLDMSLKAFDHENTSELYKEELLE	NKNPHLYRSVARDLGGTPAMINHSDMSNTI	1243
881	Qy	-----	-----	916
1244	Db	PDEKWTIYLSFLCARLIDLKELRAARL	IQTTWKYKLTDLK---RHQE---REKAAR	1297
917	Qy	VIOAVFRGHQMKYVKLFKERTQAAIILQK	FTERYLAQQLYQ-----	960
1298	Db	IIQLAVINFIAKQRLR---KRVNAALVIQ	KYWRVLAQRKLLMLKKEKLEKVNKAASL	1353
961	Qy	-----	-----	981
1354	Db	IQGYWRRYSTQRFELKLYSIILOSIRMI	IAVTSYKRYLWATVILQRHWRAVLRKQD	1413
982	Qy	QORFVELREAAIFLQ---RIWRRELFAK	LALAAETAR-----LORSQQAAS	1032
1414	Db	QORYEMLKSSLLIOSMERKWKQKMSQV	KATVILQRAFEWHRLRQAKEENSALIIQS	1473
1033	Qy	QWRYIQGRIQRHFEFLQRDLIMFVQR	MRSKWMLRQKFEQOLKAAINIQRWPAKL	1092
1474	Db	WYRMHK---ELRYQIVIRS	CVWIIQKFR---CFOAKLYKERKESIIITOKYKAYL	1525

Qy	1093	SMRKCNDADYALRSSVLKVA-----YRKATI-----QMRIDRNHHVYSIRKN	1134
Dd	1526	KGKITERTNYLQKGAAATLOQAFFRLKAHNLRCQIRAACVCIQSYYWRWRODEVRFNLKKT	1595
Qy	1135	VICLQQRLRALKMREORENVLRLRWASILLVOKRYMRQOMIODRNAYLRTKTCGINVOR	1194
Dd	1586	II----KFQAHVRKHQORQYKKMKGAAVIIOTHFRAYIFAMKVLASYQKTRSIVTLOS	1641
Qy	1195	RWRATLQMRERKNYLLQTTTKRIQIKFRA---KRE-----MKCORAE	1235
Dd	1642	AYRG---MQARXWYIHILTSVIKIOSYRAYVSKEFLSLKNATIKLOSTVKMKQTRKQ	1697
Qy	1236	FLOLKVTTLVVQKRALLQWKERQBLYHLREVTIKLO--RRFHAKQSKRPFMRKAYRG	1292
Dd	1698	YLHLRAALFTQCYSRSKCTAACKREEYQMWRSCIKLQAFVRGYLVRKQWRQLQ---	1752
Qy	1293	TQAAVSCLOWHRNHLLRKERNSFLOLROAAITLQRRYEARLNMMIKOLKSAYOLKQAAI	1352
Dd	1753	--KAVISLSQSYFR---MRKARQVYLKMVKAILVIQNYHYAKAQVNQRENFLQVKKCAAT	1806
Qy	1353	TIOTPRYAKAMQKVVLVYQOBREALIKVQRRYRGNLEMRKQIEVYQOQAVIRLOKWW	1412
Dd	1807	CLQAAVGYGVQRQ---LIQQOSTAALKIQSAFRG---YNKRVK-YQSVLQSLIIKIQRWY	1858
Qy	1413	RSIEDMWELCKAGYRIFLSSLSIQ---RKWRATVOABRQREIFLSTIRKYVLMOAFTRAT	1469
Dd	1859	RAYKTLDHTDTEHLTKAAVISISQSYRGKVRKQIRRHQ-----	1899
Qy	1470	LLMRQRRREFEMKRRAAVITQRPFRARCMLKARQDYQLIQSSVILVQRKFRANRSMKQA	1529
Dd	1900	-----AALKIQSAFR-----MAKAQKQFELFKTAALVIOQNFWATAGRKQ	1940
Qy	1530	ROBFVQLRRTAVHLQOKFRQK---RLMIEGRNCFPOLLRCSPMGPOARGFMARKRFOAL	1586
Dd	1941	CMEYIELRHVLVLIQSNMKGKTLRRLQROHKCAIIT-----QSYRVHVHVOOKWKIM	1993
Qy	1587	MTPEMMDLIHQKAAKVIOQVWRYGLIRLRKHQGL-----LDIRKRIAQ	1631
Dd	1994	-----KKAALIQIKYRAISIGREONHDLYLTCAAUVTLOSAYRGMKVRKRDKD	2042
Qy	1632	LROEAKAVNS 1641	
Dd	2043	CNKAAVTIQS 2052	
RESULT 6			
Q8N4D1	ID	Q8N4D1 PRELIMINARY; PRT; 1142 AA.	
AC	Q8N4D1:		
DT	01-OCT-2002	(TrEMBLrel. 22, Created)	
DT	01-OCT-2002	(TrEMBLrel. 22, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Similar to asp gene product.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
FN	[1]		
SEQUENCE FROM N.A.			
RP	TISSUE=Lymph;		
RC	Strausberg R.;		
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC034607; AAI34607.1; -.		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR001715; Calponin-like.		
DR	InterPro; IPR000048; IQ_region.		
DR	Pfam; PF00307; CH; 2.		
DR	Pfam; PF00612; IQ; 14.		
DR	SMART; SM00033; CH; 1.		
DR	SMART; SM00015; IQ; 11.		
DR	PROSITE; PS50021; CH; 2.		
DR	PROSITE; PS50096; IQ; 2.		



Qy	SEQUENCE	1142 AA; 133617 MW; FA5738B59F20E83A CRC64;	
	Query Match	9.1%; Score 868; DB 4; Length 1142;	
	Best Local Similarity	24.6%; Pred. No. 9.5e-48;	
	Matches	336; Conservative 217; Mismatches 492; Indels 322; Gaps 46;	
Qy	582	YLYKRLVETLKAVALFFSEOMRLPCKSVAVYVVKQALRISDRNLHLHLDVVMQRTILEL	641
Db	6	YTCRNLRLRRAACRLFTSEKMYATKLEIEARRLVVRKDRHLWKDVGERQKVLNV	65
Qy	642	LLCFNPWLRLGLEWPFCEKIQMOSNRDVLGSLFFILNRLFRN---KCEQRYSKAYTLT	698
Db	66	LLSYNPLWRLIGLETTYGELISLEDNSDVTGLAMFILNRLWNPEDIAAEYRHPTVPHLYR	125
Qy	699	BEVAETTKKSLQKILFLPLPQAKQKRIVKNPCLFVKSPKHEKTDKILLRPSSELLA	758
Db	126	DGHEALSKFTKLLLVCFDLYAKTSRLTDHDPCLFCCKDAEFKASKETILLAFSRDFLS	185
Qy	759	NIGDITRELRLGLVLOHRTFLDEPDYAFNNLAVLDLGDVRLVTRVVEILLRDLTQOL	818
Db	186	GEGLDSRLHGLGLPVNHVQTPPEFDFAVTNLAVDLQCVGLVRLVTWELLITQNDLSKKL	245
Qy	819	RVPAISRLQRFNVKLGALGALGANFQI---CGDIAAQIDVGHKRTKLSLWQLIYKF	874
Db	246	RIPAISSLQKMHNDVILQVLKSGRIELSDHGNLTLSKDIIVDRHREKTLRLWLKIAFAF	305
Qy	875	RSPKFHAAATVLQKWRHMLHVIOQRHKLWRRHRAATVIAQVFRGH---QMRKVKV	932
Db	306	Q-----VDISLNDQJKE-----EIAFLKHTSKIKTTIS	334
Qy	933	LFKTERTQAAIILQKFTRRYLAQKOLQSYHSIITIQRWRA---QQLGRHORQRFVBLRE	990
Db	335	LLSCHEPD---LINKKGRDSSGSEQYS---ENIKLLMDWNAVCAFYNNKVENFTVSFSD	390
Qy	991	AAIF-----LQRIWRRLLFAKLLAAETARLQSQQAQAASAYIQMWRTY	1037
Db	391	GRVLCYLIIHHYPCVYFPDAICQR---TTQTVECTQTGSSVNLNSSSDSSLDLSLAKF	447
Qy	1038	QLGRIQRHFLQRDLIMFVQRRMRKWSMLQKQEFQOLKRAAINIQQRWRAKLSMRKC	1097
Db	448	D-----HE--NTSELK-----ELLENKKPHLVRSAVR-----	475
Qy	1098	NADYALRSSLVKVQAVRKATIQMRIDRNHYSLURKNVICLOQRLRAIMKWRQRENYLR	1157
Db	476	--DLGGIPAMT-----NH--SDMSNTI-----PDEKVV	499
Qy	1158	LRNASILVQKRYMRQMIQDRNAYLRTKCIINVQRWRATLQMRBERKNYLHLQTTK	1217
Db	500	ITYLSFTCARLLDURKEI---RAARL---IOTTWR--KYKLTDLKRRHQREKAAR	547
Qy	1218	RIQ---IKFRAKRMKQRAEFLQKVKTVLVQK-----RRRALLQMRKERYLHLRE	1268
Db	548	IIQLAVINFLAKQRLRR-----VNAALVIQYWRVLAQKRLMLKKELEKVQNKQA	600
Qy	1269	VTI---KLORPFAOKSWRFPRKAYRGVTCQAAVSLQMWVRNH-----LLR	1310
Db	601	ASLQAMWRRYRAKVIYCKVKAACK--IQAWYRC---WRAHKEYLAILKAVKIIOGCFYT	655
Qy	1311	KRENSFLQROAAITLQRYRRAELNMIKOLKSYAOLK---QAAITITRYRAKKAMQKV	1368
Db	656	KLETRFLNVPASAIITQKWRALIPAKIAHEHFLMIKHBRAACLIQAHYRGYK---RQV	713
Qy	1369	VLYQKREAIK---VQRRYRGNLEMRKQIEVYQKQORAVTLQKWRISDRMLCKAGYR	1426
Db	714	FLRQSAALITQKIRAREAGKHERIKYIEP---KKSTVILQ-----	752
Qy	1427	RIRLSSLSIQKRWATQARRQRIEFLSTIRKVLQMAQAFIRATLLMRQORREFPKRAA	1486
Db	753	-----ALVRGWL-----RGRFLEQRAKIRLLHFTAAAYHIN-----A	786
Qy	1487	VVIQRRFRACMLKARQDYQLIOSSVILVQRKFRANRSMKQARQEFVLQRTIAVHLQOK	1546
Db	787	VRIQRAVKLYLAVNANKOV-----NSVICIOWRFAELOEKRFLOKTHSI-----	832

```

Db 258 QSSL---TFIKP-----LKTDPHPMPFAAKNMFYDERWKEQEQGFTWWNLFLT-PDD 309
QY 547 LDADLN-NKIDVCKLFNEVENKELV---APTKEQSMN-YLTKYRLETLRKADELFFS 601
Db 310 FTVKTNISEVNAATLLGIENQHKISVPRAPTKEEMSLRAYTARCLNRLRRAACRLFFS 369
QY 602 EQMRLPCSKVAVVNVKQALRISDRNLHLVDVVMQRTILELLCFNPLWLRGLGVVFGEX 661
Db 370 EKMVKAIKKLEIEIARRLLVRKDRHLWKDVGQRKQVNLWLLSYNPLWLRIGLLETYYGEL 429
QY 662 IQMQSRDVLGLSTFILNRLFRN---KCEQRYSKAYTLTEEYAEITKHSLOKILFLPLP 718
Db 430 ISLEDSDVTLGLAMFILNRLWNPDIAAEYRHPVPHLYRDGHEEALSKFTLKKLLLVLC 489
QY 719 FLDOAQKQKIVKHNPCLFVKKSPHKTKOILLRFSSELLANIGDITRELRRLGVVLOHRQ 778
Db 490 FLDYAKISRLIDHDPCLFCDAEFKASKETILLAFSRDFLSGEGDLSRHGLGLGLPVNVHVQ 549
QY 779 TFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDDLTQLRVPALSRQLQIFNVKLALGA 838
Db 550 TPDDEFDAVNLAVDLQCGVRLVRTWELLTQNWDLSSKLRIPALSRLOKMHNVDIVLQV 609
QY 839 LGEANFQL-----GGDIAAQDIVDGHREKTLISLWLIYKFR 875
Db 610 LKSRGIELSDEHGNTILSKDIVDRHREKTLRLWKTAFAPQ 650

RESULT 8
Q81ZJ8
ID Q81ZJ8 PRELIMINARY; PRT; 826 AA.
AC Q81ZJ8;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Abnormal spindle protein ASP (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Askham J.M., Springell K., Mahadevan M., Crow Y.J., Markham A.F.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AY099891; AAM44120.1; -.
DR InterPro; IPR001715; Calponin-like.
DR Pfam; PF00307; CH; 2.
DR PROSITE; PS50021; CH; 2.
FT NON_TER 1
FT NON_TER 826
FT SEQUENCE 826 AA; 94875 MW; 970ED55486C4C88B CRC64;

Query Match
Best Local Similarity 34.1%; Pred. No. 3.6e-39; Length 826;
Matches 198; Conservative 104; Mismatches 210; Indels 69; Gaps 18;

QY 334 QEQIQI-----KARD-LNSSSSSSASLAGQOEFILFNHSE-----ILAQSSRF 375
Db 100 RENQTEINKPKACRLNSA-----VGEHKYNNQKEDPHSYLPIDPILSKSKY 152
QY 376 NLHEVRKSVKSPVKNPKHRSHELFSFDAPSNESLYNETVAISPPKQKQVEDTTLPR 435
Db 153 K-NEVTPSSTTASVA---RKRKSDGSMEDA-----NVRVAITEHTVEVRIKHS 199
QY 436 SAAPANASARSASAHAWPQAOKKFLAQTMSLMKPP-----ATPRKVRTDTSQPSVKLY 490
Db 200 PSEPKTSVAKKTKNVWTP--ISKRSINRKNLKKKTDLSIFRTSKTKTKTPIIAYA 257
QY 491 DSELYMQTCINPDPFAATTIID-----PFLASTMYLDEQVDRHQADFKKWLNALVSIAD 546

```

```

Db 258 QSSL---TFIKP-----LKTDPHPMPFAAKNMFYDERWKEQEQGFTWWNLFLT-PDD 309
QY 547 LDADLN-NKIDVCKLFNEVENKELV---APTKEQSMN-YLTKYRLETLRKADELFFS 601
Db 310 FTVKTNISEVNAATLLGIENQHKISVPRAPTKEEMSLRAYTARCLNRLRRAACRLFFS 369
QY 602 EQMRLPCSKVAVVNVKQALRISDRNLHLVDVVMQRTILELLCFNPLWLRGLGVVFGEX 661
Db 370 EKMVKAIKKLEIEIARRLLVRKDRHLWKDVGQRKQVNLWLLSYNPLWLRIGLLETYYGEL 429
QY 662 IQMQSRDVLGLSTFILNRLFRN---KCEQRYSKAYTLTEEYAEITKHSLOKILFLPLP 718
Db 430 ISLEDSDVTLGLAMFILNRLWNPDIAAEYRHPVPHLYRDGHEEALSKFTLKKLLLVLC 489
QY 719 FLDOAQKQKIVKHNPCLFVKKSPHKTKOILLRFSSELLANIGDITRELRRLGVVLOHRQ 778
Db 490 FLDYAKISRLIDHDPCLFCDAEFKASKETILLAFSRDFLSGEGDLSRHGLGLGLPVNVHVQ 549
QY 779 TFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDDLTQLRVPALSRQLQIFNVKLALGA 838
Db 550 TPDDEFDAVNLAVDLQCGVRLVRTWELLTQNWDLSSKLRIPALSRLOKMHNVDIVLQV 609
QY 839 LGEANFQL-----GGDIAAQDIVDGHREKTLISLWLIYKFR 875
Db 610 LKSRGIELSDEHGNTILSKDIVDRHREKTLRLWKTAFAPQ 650

RESULT 9
Q81ZJ9
ID Q81ZJ9 PRELIMINARY; PRT; 882 AA.
AC Q81ZJ9;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Abnormal spindle protein ASP (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Askham J.M., Springell K., Mahadevan M., Crow Y.J., Markham A.F.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AY099890; AAM44119.1; -.
DR InterPro; IPR000048; IQ region.
DR Pfam; PF00612; IQ; 34.
DR SMART; SM00015; IQ; 33.
DR PROSITE; PS50096; IQ; 2.
FT NON_TER 1
FT NON_TER 882
FT SEQUENCE 882 AA; 107102 MW; 84AD920618B6A1F6 CRC64;

Query Match
Best Local Similarity 7.2%; Score 688; DB 4; Length 882;
Matches 225; Conservative 173; Mismatches 314; Indels 206; Gaps 28;

QY 868 WLIYKFRSPKFAAATVLQKWRH-----WLH-----VITQRRIR-----HKELMRRHR 913
Db 6 WHL---RQAKEENSAIIQSWYRMHKLRYIYRSCVVIQKFRFCFQAOKLYKRRKE 62
QY 914 AATVIAQVGRGHQMKYVKLFKTERTAIILOKTRRYLAQKQYQSHSIITLQWRW 973
Db 63 SILTIQYKAYLKGKIERNTYLRQKRAAIQQAQAFRR-LKAHNLQRTRACVCIQSYWR 121
QY 974 AQQ-----LGRQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAF 1014
Db 122 MEQDRVRFNLKTTIKFOAHYKHOQKQYKMKKAAVITHTHFAIFAMKVLASYQK 181

```

QY	1015	AR	-----LQRSQK-QOAAASYIQ-----MOWRTYQIGRIQRBHEFLRQDLIMFVQRRMS	1063
Db	182	TRSAVILQSAVGMQARKMYIHILTSVIKIQSYRAYVSKFEFLSKNATI----	KLQS	237
QY	1064	KWSMLEORKEFOOLKRAINIQORWRAKUSMRKCNADYALALRSVLKVQY-----	1114	
Db	238	TVRMKOTRKQYLHLRAAALPQOCYRSKKTAAOKRBEYMQRBESCICLOAFVGVYLVVRKQ	297	
QY	1115	-----RKATI-----QWRIDENHYISLRKNVICLOQLRAIMQWORENYELRNASIL	1164	
Db	298	MRLQKRAVTSLOSVERNKRARQYLLKMYKAILVIQNYHAYKAOVMQKNFLQVKAATC	357	
QY	1165	VQKRYR-----MRQOMI-----	QD	1178
Db	358	LQRAYRGYKVRQLIKQOSIAALKIQSAFRGYNKVKYQSVLQSIILKIQRWYRAYKTLHDT	417	
QY	1179	RNAYLFRKCIINVQ-----RRWRATLQMRERKNVYHLQTTTKGIOIKFRAKREMKQRAE	1235	
Db	418	ETHFLTKAAVTSLOSAYRGWKVRKQIRREHQAAL-----KIQSAFR-----MAKAQKQ	466	
QY	1236	FLQLKKVTLVQKERRALLQMRKERQBYLHREVITIKLO-----RPHAKS-----	1282	
Db	467	FRLEFTAALVIQONFRATWATRGRQCMEYIELRHAVLQSMWKGKTLRLOLQOHKCAII	526	
QY	1283	-----MRFWAKYRGTHAAVSCLOMWRNHLAKRENSFLOLRQAATILORRYEARLN	1336	
Db	527	IQSYRMHVQOKKWMKKGALLIOKTRYAISIGREONHLYLTKAAVTVLOSAVIG----	583	
QY	1337	MIKOLSYAOLQOAAITIQTRYAKKAMQOVVLYQKOREAIIKVORRYRGNLEMRKQIE	1396	
Db	584	-MKVRGEIKDCNKAAVTIQSKYRAYTKKK-----YATYRASAILIQWYRGIKITNHQHK	638	
QY	1397	VYQKQOAVIRLOKWSIR-----DMRLCKAGYRRIRLSSUSIQOR	1437	
Db	639	EYLNKKTALIKIQSVYRGIRVRRHIQHMRAAATFIKAMFQKHQSRISSYHTRKAAAVIQV	698	
QY	1438	KWRATVOARQOREIFLSTIRKVRLOMAFIATILLMQOOREFEMKRAAAVVIQRRFARC	1497	
Db	699	RCRAYQGWQREKYLITLKAVKVLOASFRGV-----RVRETLRKMQTAATLIQSNYR----	751	
QY	1498	AMLKAROD--YOLIQSSVILVORFNRANRSMKQARQEFVOLRTTAVHLOKQFKGKILMIE	1555	
Db	752	----RYROQTFFNKUKTKTVQOQRYWAKERNTOFYKNLKRHSVIVIOAIFGKK-----	804	
QY	1556	QRNCFOLLRSGMPGOARAGFMARKEFOALMTPEMMDLIRQKRAKVIORYWRGYLIRR	1615	
Db	805	ARRHLKMHHTAATLIQRRFRTLWMRRRFLSI-----KKTAILIQKYRAHLCTK	853	
QY	1616	ROKHQGLLIDTRKTAQUR	1633	
Db	854	H--HLQFLQVONAVIKIQ	869	

## RESULT 10

Q9NSV1	_L0
ID	Q9NSV1
AC	PRELIMINARY; PRT; 726 AA.
DT	O9NSV1;
DR	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein FLJ10549.
OS	Homo sapiens (Human).
OC	Homo apitans (Homoan).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Iscgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA	Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA	Tanase T., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K.,
RA	Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA	wakamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.;

  

QY	1527	KQARQREVFOLRTIAVHLQQKFRGCKRLMIQRNCFQLLRCSMPGFQARAGFMARKRFOAL	1581
DB	407	KRTIQKHSI-----KKIEHQECLSQENRAASVIQKAVRHFLRK----	449
QY	1587	MTPEMDLIHQRAAKV--IQRVWRGLIRRRQKHOGLLDIRKRIAQLOREQAKAVNSVC	1644
DB	450	-----QEKFSTGIKIQAALRWGRYSWNRKNDCTKIKAIRLSUQVYNVIREENKLYK	500
QY	1645	KVOAEVRFLRGRTIASDALAVLSQDLRLSRTPVPHLLMWCSEFMSTFCYGIMAQAIRSEVD	1704
DB	501	RTALALHYLLTYKHLSAILEALKXHELVVTR-----LSPLCCENNAQSAGAIISKI	548
QY	1705	KLIERCSSR-----ILNLARYNSTVTNTFQBGGLVTTAQMLL-----	1742
DB	549	FVLIRSCNRSLPCWEVIRYAVQVLLNVSKYEKTTSAYDVENCIDLLELLQIVREKPGN	608
QY	1743	RWCDKOSEIFNTICTLIHWFAHCPEKRKKIKIHVDYWNDEALYMWRETKKILVARKEKMQNA	1802

Db 609 KVADKGGSIPTKTCCLAILL---KTTNRASDVRSKVVDRISYLYKLTAHKHKM--NT 663

QY 1803 RXPPPMTSGRYSKQ-----INFTPCS-----LPSLEDPFGIIRYSPTFFISSVAFDT 1851

Db 664 ERI-----LYKQKNSISIPFIPVTRIVSRKLPDVLRRDNMBEETNPLOAIQM 717

QY 1852 ILCKLQI 1858

Db 718 VMDTLGI 724

RESULT 11

Q84228 PRELIMINARY; PRT; 1110 AA.

AC Q84228;

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE OSJNBA0036M16.11 protein.

GN OSJNBA0036M16.11.

OS Oryza sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI\_TaxID=39947;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Sasaki T., Matsumoto T., Katayose Y.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC

RT clone:OSJNBA0036M16.";

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP005103; BAC56022.1; -.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR000225; Armadillo.

DR InterPro; IPR001715; Calponin-like.

DR InterPro; IPR000048; IQ\_region.

DR Pfam; PF00307; CH; 1.

DR Pfam; PF00612; IQ; 5.

DR SMART; SM00185; ARM; 1.

DR SMART; SM00033; CH; 1.

DR SMART; SM00015; IQ; 4.

DR PROSITE; PS50021; CH; 1.

DR PROSITE; PS50096; IQ; 2.

SQ SEQUENCE 1110 AA; 125271 MW; 830CB31B3A57958D CRC64;

Query Match 3.9%; Score 370.5; DB 10; Length 1110;

Best Local Similarity 19.3%; Pred. No. 2.8e-15;

Matches 268; Conservative 212; Mismatches 487; Indels 425; Gaps 52;

QY 419 AISPCKQVEDTTLPRSAAPANASARSSAHAWPHAQSKFKLAQTMSLMKKPAIPKPV 478

Db 7 AVSSPPFD-LSNURTPNPRAPPNPKSSAKKEPLP-----SATPTTR 47

QY 479 RDTSIQSVKLYDSELYMQTCINPDPAATTTIDPFLASTMYLDEQ-----AV 526

Db 48 RRGGPPPP-----PRCAATATPLARRLRLDLQSSARRAESGRDGL 92

QY 527 DRHQADPKKWLNALVIPA-----DLDA-DLNKIDV-----GKLFNE 563

Db 93 RAFASATSLSLLLRDPSCAGCAPSAAAARVTRDAPAHGVQGRDAVGERARGSPKR 152

QY 564 VRNKLVAAPTKEQSNYLTLYEL-ETLRKA-----AVELPFEQMLPCKSVAVY--- 614

Db 153 HRGEGDGGGPPRRKTKMTPMAASLRDSLRVCSLDVTVTMRGSHMSREACEEVLVMQCQ 212

QY 615 ----VNKQALIRIRSDRNLDHVVWQRTILELLLCFNPLRLGLEVVFG-----EKIQW 665

Db 213 ICKNIINGRLKMEHCPELSDLRDKATPIFCWYFPKWLRIGLHVLGDSWIQNESQK 272

QY 666 SNRDIVLSTFIILNRLF-----RNKCEQRYSKAYLTETEYAEFTIKKHSLOKILFL 716

Db 273 KDEVAFLKVLKQLFVHMNTQPSALNKAPEGHRAC-----YAEAAASNLKELFL 326

QY 717 LPFLDOAKKR-----IVKHNPCLFVKSPKHKTKOILLRFSSELLANIGDITRE 766

Db 327 VVALDRAKIESGLPSBSGIDGLDGGSPFLFCRTEIKSSRQIVQESLGEVHGEGDILMH 386

QY 767 LRLGLVLOHROTFDLDFDYAFNNLAVDLRDGVRLLTRVVEVILLRDDLTQLRV--PAIS 824

Db 387 LNINGYKLYNQALASEYDFTVGNLFEDLDQGIILCRIIQ--LLTSDASILKVIAPSDT 444

QY 825 RLQRIFNVKLALGALGEANFQ-----GGDIAQDIDVGHREKTLISLLWLIYKFRSPKFH 880

Db 445 YKKRLHNTMAIQYIKQAGFPLSDADGLSIAEDIVNGDKELILALLWNFIYMLPVLV 504

QY 881 AAATVLOKWWRRHVLHVQRRIRHKELMRHRAATVIAVFRGHQMRKVVVLFKTERQT 940

Db 505 NETSVAQ-----EISRLK 517

QY 941 AAAILQKFTRRYLAQKLYQSVHSIITIQRWRAQQLGRQRQFVELREAAIFLQRIWR 1000

Db 518 APVSEQSISEMKSQTGLLYD-----WI 539

QY 1001 RLIFAKKLLAAETARLQSQKQAAAASYIQMWRTYQLGRQORHEFLQRDLIMFVQRR 1060

Db 540 QVVCAYKGISVESSSQIDRRALNYFISYILININIPNPL-----578

QY 1061 MRSKWSMLRQKEFQQLKRAAINIQORWRAKLSMRKCNADYLALRSSVLKVQAYRKATIQ 1120

Db 579 ---KETLSDCKRELFSCHTDMADIITYQFNIGKDLPGWNILANDVLFE--KSAAIL 633

QY 1121 MRIDRNHYYSLRKNVITCLOQLRLRAIMKREORENYL--RLRNASILVQKRYMRQMIQD 1178

Db 634 LAFSLSHLNVRR-----LEQLKNLIDSKLDHQSIVTE-----666

QY 1179 RNAYLTRKCIINVQRWRATIQMRERKNYHLQTTTKRIQIKFRAKRMKQRAEFLO 1238

Db 667 -----VSPRRSRGTTDMK-----CHFPQTEE--TDGSRSTREW-----698

QY 1239 LKKTVLVOKRRRALLQMKERQEVYHLREVTIKLQRRFHAQ-----KSMRFMRKYRG 1293

Db 699 ---AATVIQTQARRLNAMSK---YCKLNATQPCNKGHDPASSPLKSIADSSCDSAT 751

QY 1294 Q-----AAVSCLOMWRNHLKRRRNSFLQLRQ--AAITLQRRYRARLINMIKOLK 1342

Db 752 KLVCEDDVDCSSNCQVLYHDPVSTKVD---FLFCEKAWAARKIQFAYRRFAHRISRI 808

QY 1343 STAQLKQAATITQT-----RYRAKAMQKVVLVYQKOREAIIKVORRYGNLEMRKQI 1395

Db 809 S-----AAIKIQSHWRCSFVIRFKRQIQNTTIOAVARTSAISVQSFVRGWL-IRKQV 861

QY 1396 EYVQKQOQAVIRLOKWRISIRDMRLCKAGVRRIRLSLSIQ--RKWRATVQARROREIF 1452

Db 862 ---KQILCSVYLQIRWVRQVLFESKR-----SVIVIOAHVRGWIARQTVRNK--- 908

QY 1453 LSTIRKVRMLQAFIRATILMQORREFEMKRAAVVIORRFRACAMLKARQDYQIQQSS 1512

Db 909 ----KXITIIQSVKAYLLRKSQEIITDDIMHMI---QKLRVNC-----FRLVSPG 953

QY 1513 VILVQKFRANKSMQARQEFVQLRTIAVHLQOKFRGKLMITEORNCFOILLRCSMPGQA 1572

Db 954 T-----ATEHSEKCCOTIVNAGAVEILLQINLLNRGVPDQEVILKQVL----- 996

QY 1573 RARGEMARK--RFOALM-----TPEMMDLIQKRAAKVQIRYWRGYLIRRKQHQGLL-- 1623

Db 997 ----FTLRINARPNLQPVIANTPQAVEIVFOELL-----RSTGEFFVA 1037

QY 1624 -DIRKEIAQLRQEAKAVNSVR-----CKVQEAVERFL-----RGRF-IASDALAVLSOL 1669

Db 1038 CHILKELCESSEGHKIPARALKHIRLGLCVQLEKKVDLDKNGTGVAKD-----NNL 1092

QY 1670 DRLSR--TVPHL 1679

Db 1093 RRLGEAVTLHL 1104



[illegible]







*This Page Blank (uspto)*



```

Db 648 RRHLLKSEGEERHRRHQLRRECEQERREQLRKREERLEQLKREHE-----EERRE 701
Qy 951 RYLAQKOLYQSYHSIIT-IQW-WR-----AQQLG----- 978
Db 702 QELAEEOQARERIKSRIPKQWOLESEADARQSVLLEAPQAGRAEAPQOEKEKRE 761
Qy 979 -----RHRQRQFVELREAAIFLQRIW-----RRLPFAKLLAAAEATLQRS 1020
Db 762 SELQWEEERAHROQOEERDRDFTWQ--WOAEKSEGRORLSARPPLRQERQRLRAE 819
Qy 1021 OKQAAASYIQWQRTYQLGRIQRH-----EFLQRDLIMFVQRRMRKWSMLQ--RKE 1073
Db 820 ERQOEQRFLEPEEKEQGRQRREREKELQFLEEEQL---QRRERAQQLQEEEDGLQE 876
Qy 1074 FQQLKRAAINIQORWRAKLSMRKNADYALRSSVLKVOAYRKATIOMRIDNHYSLRK 1133
Db 877 DQERRRQQRDQKRWOLEBERRRH-----TLYAKPALQEQ-----LRK 918
Qy 1134 NVICLQQLRAIMKMRQENYLRNASILVOKYRMRQOMIODENAYLRTKCIINVQ 1193
Db 919 EQQLQEEBEL--QREERKRRQEQ-----EROYREEEQQLQEEEQLLREER---EKR 968
Qy 1194 RRWRATLQMRERKNYLHLOTTTKRI-----QIKFRAKREMKQRAEFLOLKKVTLVVQK 1248
Db 969 RQERERQYRKDKK---LQKEQLGEEPEKRRQREKKYREEBELQEEEQLLREE 1024
Qy 1249 RRRALLQMRKEROEYHLREVITIKLORFPAKSMRFRKRVGTQAAVSCLOMHWNRHL 1308
Db 1025 RE-----KRRQEW--EROYRKDELOQEEBELREBERRR-----LQERERQY- 1067
Qy 1309 LRKERNFLOLQAAITLQRYRARNLMIKLSYALQKAAITI-----QTRYRAKKA 1363
Db 1068 ---REBELQEEQLGEERTRRRRQELRQRYRKEBELQEEEQLLREPEKRRQERE 1124

```

RESULT 2

```

US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fredrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-644-94

```

Query Match 2.9%; Score 277; DB 2; Length 1898;

Best Local Similarity 21.7%; Pred. No. 5.6e-14; Indels 224; Gaps 44;  
Matches 188; Conservative 155; Mismatches 301;

```

Qy 891 RRHLLVVIQRIHKKELMRRHRAATVIAVFRGHQMKYVVKLFKTEQTAAILQKPTR 950
Db 648 RRHLLKSEGEERHRRHQLRRECEQERREQLRKREERLEQLKREHE-----EERRE 701
Qy 951 RYLAQKOLYQSYHSIIT-IQW-WR-----AQQLG----- 978
Db 702 QELAEEOQARERIKSRIPKQWOLESEADARQSVLLEAPQAGRAEAPQOEKEKRE 761
Qy 979 -----RHRQRQFVELREAAIFLQRIW-----RRLPFAKLLAAAEATLQRS 1020
Db 762 SELQWEEERAHROQOEERDRDFTWQ--WOAEKSEGRORLSARPPLRQERQRLRAE 819
Qy 1021 OKQAAASYIQWQRTYQLGRIQRH-----EFLQRDLIMFVQRRMRKWSMLQ--RKE 1073
Db 820 ERQOEQRFLEPEEKEQGRQRREREKELQFLEEEQL---QRRERAQQLQEEEDGLQE 876
Qy 1074 FQQLKRAAINIQORWRAKLSMRKNADYALRSSVLKVOAYRKATIOMRIDNHYSLRK 1133
Db 877 DQERRRQQRDQKRWOLEBERRRH-----TLYAKPALQEQ-----LRK 918
Qy 1134 NVICLQQLRAIMKMRQENYLRNASILVOKYRMRQOMIODENAYLRTKCIINVQ 1193
Db 919 EQQLQEEBEL--QREERKRRQEQ-----EROYREEEQQLQEEEQLLREER---EKR 968
Qy 1194 RRWRATLQMRERKNYLHLOTTTKRI-----QIKFRAKREMKQRAEFLOLKKVTLVVQK 1248
Db 969 RQERERQYRKDKK---LQKEQLGEEPEKRRQREKKYREEBELQEEEQLLREE 1024
Qy 1249 RRRALLQMRKEROEYHLREVITIKLORFPAKSMRFRKRVGTQAAVSCLOMHWNRHL 1308
Db 1025 RE-----KRRQEW--EROYRKDELOQEEBELREBERRR-----LQERERQY- 1067
Qy 1309 LRKERNFLOLQAAITLQRYRARNLMIKLSYALQKAAITI-----QTRYRAKKA 1363
Db 1068 ---REBELQEEQLGEERTRRRRQELRQRYRKEBELQEEEQLLREPEKRRQERE 1124

```



QY 1208 ---NYLHLOTTTKRIQIIPRAKREMKQORAEFFLOLKKVTVVQKRRAALLQMRKERO--B 1262  
Db 1572 ELTN--RLQBSQBEIQIMKEEMKQVQ--BALQIERDQL--KENTKEIVAKMKESEKE 1626  
QY 1263 YLHREVTIK---LQRPFAQK---SMRFRAKYRGTOAAVSCLOM 1302  
Db 1627 YQFLKTAVNTEQKCEIHLKEQFQTKLNLENIETENIRLTQILHENLEEMRSV--- 1683  
QY 1303 HWRNHLKREPNFLOLRQAATL---QRRYRANIMIKOLKSYAQLK 1348  
Db 1684 ---TKERD---DLRSVEETLKVREDQLENLRETTITRDLEKOBELKIVHMLKEH 1732  
QY 1349 QAAI---TIQTYRAKAKOKVVIYQKREAI---IKVORRYR--GNLEMKOLEVY 1398  
Db 1733 QETIDKLRGIVSEKTEISNMOKDL---EHSNDALKAQDLKIQBELRIAHMLKEQOETI 1789  
QY 1399 QKORQAV---IRLOKWRGIRDMRLCKAGYRIRLSLSIQRKWR 1440  
Db 1790 DKLGIYSEKTKLSNMOKDLENSNAKLOE---KIQELKANEHQILITLKDVNETQKVS 1846  
QY 1441 ATVOARQREIFLSTIRKVRIMQAFIRATILMROQRREFEMKRRAAVVIQRRFRACML 1500  
Db 1847 EMEQLKQIKDQSILTSKLEIENLNLAQELHENLEEMKSVMKERDNL---RRVE---ETL 1900  
QY 1501 KARODYOLI QSSVILVQKFRANRSMKQARQEFVQLRTIAVHLOQKFRGKELMIE--QRN 1558  
Db 1901 KLERD-QIKESLQETKARDLBIQBELKTARMLSEKHEKTVDKLREKISEKTIQISDIQKD 1959  
QY 1559 CFQLLRCSMPQFQARARFARKFQALMTPEMDLIRQK--RAAKVIQRYWRGYLIR-- 1614  
Db 1960 ---LDSKDELQKIQE--LOKKELOLLRVKEDVNMHSHKKNEMEQKKQPEPNVLKCE 2014  
QY 1615 ---RQKHQGLDITRKRIAQURQEKAVNSVRCKVQBAVRLGRFIASDALAVLSQ 1668  
Db 2015 MDNFQTKLHESLEIR-IVAKERDEL---RIKESLKMERDQFIAT--LREMLA 2064  
QY 1669 LDR-----LSRTVPHLMWCSEFMFCYGINAQAIRSEVDKQLIRCSRIILN 1717  
Db 2065 RDRONHQQVPEKRLISDCQOHLN-----BSLREKCSRIKEL 2100  
QY 1718 LARVNSTVTNTFOEGGLVTIAQMLRLWCDDKSEIFNTLCTLIWVFAHCPKRRKIHDYMT 1777  
Db 2101 LKRYSE-----MDHYECLNRLSLD--LEKEIEBPHRIMKLYVLISYVTKIEQHECIN 2153  
QY 1778 NPEAIYM--VRETKKLVARKEMKQNAKPPMTSGRYKSOKIN-----PTP 1822  
Db 2154 KFEMDFIDEVEKQKELLIKIQHLOQDCDVP---SRELRLKLNQNDMLHIEITLKDFSE 2209  
QY 1823 CSLPSLEPDF 1832  
Db 2210 SEFPSIKTEP 2219

## RESULT 4

US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

## Query Match 2.8%; Score 268; DB 1; Length 2482;

Best Local Similarity 18.0%; Pred. No. 5e-13;  
Matches 299; Conservative 294; Mismatches 638; Indels 428; Gaps 69;

QY 105 HPPQAPLVEKNVYK-----TPQEFVVISPPRSLKENLSMTP--GNLL 148  
Db 745 HPVSLAPLDESNSYEHLLTSLDSKEYMHFAELQKFLSQSEHKILHDOHCQMSKMSLQ 804  
QY 149 DVIDNLRTPLITETRGGOATIPPDNLAAWPTPLK-----GNVKSANDMRPRIT 200  
Db 805 TYVDSL-----KAENLVLTNLFNFGDLVKEMQGLGELGVLPSLSSSCVPS 854  
QY 201 PDDLLED---QPATNKTFDVKHSETI-----NISLDTLDCSRIDGQPHPLANKTTIV 249  
Db 855 LSSLDGSSFYRALLQIGDMSELLNLEGAVSANQCSVDVFCSSL----- 899  
QY 250 HATITRALACHHEGSPPTPTKSAIHLKRIKLVGSLPKRYSESMKDLSSLSPQTK 309  
Db 900 -----QENLTKRKTSPAPA-----KGVEELESCEVYRQSLKLE----BEK 937  
QY 310 YAIQSGMENLNMKIRSTEQ-----NRYQBOQIQIKAKDLNLSSSSSSEASLAG 357  
Db 938 MESQIMKN-----KETQLEQLLSERQELDCLRKQYLSENEQWQOKLTSVILEMESKLA 994  
QY 358 QOEFNFHSEILAQSSRFNLH--FYGRKSVKSGPVKNPKRSHLSFSDAPSNESLYRNE 416  
Db 995 EKK-----QTEQLSLEVARLQLOGLDLS-----RSLIGDITDAIQG-----RNE 1037  
QY 417 TVAISPPKQKQVEDTTLPRSAAPANASARSSAHAWPHAQSKFKLAQTMIMKPP---A 473  
Db 1038 SCDIS---KEHTSETT---ERTPKDHVHIQCDKQAQDLNLDIEKITETGAL--KPTGEC 1089  
QY 474 TPKVRDTSIQPSVKLYDSLYMOTCINPDPAATTIDP--FLASTMYLDEQAVDRHQA 531  
Db 1090 SGEQSPDTNYEPPE--DKTQGSSECTISELFSGPNALVPMDFLG-----NQEDIHNLQL 1142  
QY 532 DFKKWLN---ALVSI PADLADLANKKIDVGLKFNVRNKLIVAPTEEQSNVYTKYRL 588  
Db 1143 RVKETSNNENRLLHVIEDRDR-----KVESILNEM-----KELDSLHLQEVQL 1186  
QY 589 ETLKAAVEL-----FSEQMR-LPCSKVAVYVKNQALR--LRSDNHLHDVVM 634  
Db 1187 MTKIEACIELEKIVGELKKENSIDLSEKLEVPSCDQELLQVETSEGNLSLEMAADKSS 1246  
QY 635 QRTILELLLCFNPLWRLGLEV-----VFGEKIQMSQNRDIVGLSTFILNRL 681  
Db 1247 REDIGDNVAKYNDKWKERFLDVENELSRIRSEKASIEHEALYLEADLEVQTEKLCLEKD 1306  
QY 682 FRNK-----CBEQR-----YSKAYTITTEYAEATIKKHSLOKILFLFPF 719

Db 1307 NENKQVIVCLBEELS SVT SERNQLRGELDTMSKTTTALDQSEKMKKE-----1354  
QY 720 LQQAQKRVKPNP-----CLFVVKSPHKYKIDILLRFPSSLLANIGDITRELRLGVVLO 775  
Db 1355 -----KTOELESQSECHICQIAEAEVKEKTELLQTLSSDVSLDKDTH-----LQ 1402  
QY 776 HROTFLEDFYAPNNLAVDLRDCGVLRVTRVVEVILLRDDLTRQLRVPALSRLORIENVKLA 835  
Db 1403 EKLOLEKDSQALSLLTKCELENGIAQLNKEKELLVKESESLOARL-SESDYKEL-NVSKA 1460  
QY 836 LGAL-----GRANPOLGGDIAAQDIPVGHREKTLSSLWLIYKFRSPKPHAAATVLOKWMR 891  
Db 1461 LEAALVEKEGFAIRLS-----STQBEVHQLR-RGIEKLRVRIEADKKQLHIAEKLERERE 1516  
QY 892 RHWLHVVIQRRIRHKELMRHRAATVIOAVFRGHQMKRYVKLTKTERTQAAITLOKPTRR 951  
Db 1517 NLSLKDKNLEKLELQSENGEQLVILDA-----ENSKAEVETLKTQTEEMARSLKVPFEL- 1571  
QY 952 YLAQKOLYQSYHSIITIORWRAQOLGRQHRQRFVELREAAIFLQRIWRRRLFAKLLAA 1011  
Db 1572 -----DLVTIRS--EKENLTQIOEKQGLSELDKLLSSF-----KSLLEE 1610  
QY 1012 AETARLQSQKQAAASYTQMOWRTYOLGRI-----ORHEFL 1048  
Db 1611 KEQAEIQIKEESTAVEMLQNLK--ELNEAVALCGDQIEIMKATEBOSLDPPITEEHOL- 1667  
QY 1049 RORDLIMFQRM-----RSKWSMLEQKBPQ-----LKRAAINIQORWRAKLSMRKNAD 1100  
Db 1668 ---RNSIEKURALEADEKKOLCVLOQKSEHHDLLKGRVENL-----ELEIARTNOE 1721  
QY 1101 YLALRSSVLK-----VQYRKATIOMRIDRNHYSLRKNVICLOQLRAIMK 1147  
Db 1722 HAALRAENSKGEVETLAKIEGTQSLRGLELDVVTIRSEKENLTNELQEKERISELEI 1781  
QY 1148 MREQRNYLRLNASILVQKRYMRQ-----QMIQDRNAYLTRKCI--NVQRRWRATL 1200  
Db 1782 INSSPENILOEKE-----QEKVQMEKSSSTAMBLQTLKELNERVAALHNDQEAACKAE 1836  
QY 1201 Q-----MRRERKNYILHQTYYK-RYQIKFRAGREMKQBAEFLOLK 1240  
Db 1837 QNLSQVECLELEKALQLOGLDEAKNNYIVLOSSVNGLIQEVEDGKQLEKDEERISLKL 1896  
QY 1241 KVTLVVQKERRALLQMKERQ-----EYLHLREVTIKLQRRFHAQKSMRFRKRYGTQAA 1296  
Db 1897 NQIQDEQLVSKLSQVEGHEHQLKEQNLRLNLTVLEBK-----IQVLQSKNASLQDT 1950  
QY 1297 VSLQMHWRN-----HLLRKERNFSL-----QLRQAAITLQRRYAR 1334  
Db 1951 LEVLQSSYKRNLELELTAKMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEBLSGE 2010  
QY 1335 LN-----MIKOLKSYA-OLQQAATITQYRAKAMQK-OVVLVYQKOREALIKVQRR 1384  
Db 2011 KURLAGELQLLEIEIKSSKDDOLKBLTENSELCKSLDCMHQDQVEKEGVREIEAYQLR 2070  
QY 1385 YRGNLEMRKQIEVYQORQAVIRLOKWRMSIRDMLCKAGYR-----RIRLSLSIQKWR 1440  
Db 2071 LH-EAEKHQALLDNTNKQYVEIQYREKLSKECELSQKLEIDLILSSKEELNNSLK 2129  
QY 1441 AVQARRQREIFLSTIRKVRMLMOAFIRATILMRQORREFEMKRAAVVIOQRFRACML 1500  
Db 2130 ATTO-----ILEELKTKMDNLKYNVQNLKXENERAQCKMK-----LLIK-----SKQL 2173  
QY 1501 KARQDQLIOSSVILVORFRANRSMKQARQFVQLRTI 1539  
Db 2174 E--EKEILOKEL-----SOLQAQAEKQKGTGTV 2199

## RESULT 5

US-09-976-594-736

; Sequence 736, Application US/09976594

; Patent No. 6673549

; GENERAL INFORMATION:

; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 736  
; LENGTH: 1530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Inocyte ID No. 6673549 3151579CD1  
US-09-976-594-736

Query Match 2.8%; Score 264; DB 4; Length 1530;

Best Local Similarity 18.5%; Pred. No. 5.1e-13;

Matches 294; Conservative 277; Mismatches 517; Indels 498; Gaps 69;

QY 263 EGGSPSP-----RTPTKSAIHLKR-----DIKLVGSPLRKYSEGMKDL 301  
Db 68 QNGTPTFPDPSSLDPTTSPVGPDPASPGVAGPHDNLKRSQGTSAEGSVRKEALQSLR-L 126  
QY 302 SLLSPOTKYAIGS--MPNLENMKIRSIQONRYVYRQOIQIKAKDLNSSSSEA-----353  
Db 127 SLFMQETQLCSTDSPLEKEQVR--LOAKMLEQLKQVRVKRQQRSSQATKTRLFS 185  
QY 354 -----SLAQOQEFLEHSEI-----LAOSSRENLHEVGKSVKSPVKN 392  
Db 186 TLDPELMNPEMLPRASTLANTKEYSFLRTSPRGPKVGLPLPAHPREKTKSRKSKIS 245  
QY 393 PHKRSHELFSFDPASN-----ESLYRNET-----VAISPPKQORVEDTTLP-R 435  
Db 246 LADYRTEDSNAGSGNVFAPDSTKGLKQNRSSAASVSEISLSDTDRLDRLENTSAGD 305  
QY 436 SAAPANASARSSAHAWPHAQSKFKLAQMSLMKPAKPRKVRDTSIQPSVKLYDSELY 495  
Db 306 SVSEVDGNDSDSYSSASTRGTYGILSKTVGTQDTP-----Y 343  
QY 496 MQTCINPDPFAATTIDPPLASTWYLDQAVDBHQADPKKWLNALVSIADLDADLNKKI 555  
Db 344 M--VNGQEIADT-----LGQFESIKDVLOQAAAAEHQ 373  
QY 556 DVGKLFN-EVRNK-----ELVVAPTKEEQSMNYLTKYRLETLRKAAVELPFSEOM 604  
Db 374 DQGEVNGEVRSEKSDISCVSLESAAETQEM-----LQVLK-----EKM 415  
QY 605 RLPCKSVAVYVN-KOALRIRSDRNHLHDVVMORTILELLCFNPMLRLGLEVVVFEKIQ 663  
Db 416 RLEGQLEALSLEASQALK-----EKA 437  
QY 664 MOSNRDIVGLSTFILNLFNFKCEQRYSKATLTTEEYAEATIKKH--SIQKILFLPLFD 721  
Db 438 LQA--QLAALSFKQAQV---ECSSHSSQORQOSLSSE-VDTLKQSCWDLERAMTDLQNNML 491  
QY 722 QAKQKRIKVNKPCFLFKKSPHKETKDILLRFSSELLANIGDITRELRLGYVLRHQRTFL 781  
Db 492 EAKNASLASSNNDLQVAREQYQ-----RLMAKVEDMQRM-----526  
QY 782 DEFDYAFNNLAVDLRSGVRL--TRVVEVILLBDDITRQLRV--PAISRLORI---FNVKL 834  
Db 527 -----LSKDNVTVDLRQOMTALQSQQLERLTTLTSKLKSAQABISSLQSVRQWYQQQL 582  
QY 835 ALGALGEANFOLGGDIA-----AQDIVDGHREKTLISLLWOLI-YKFRSPK-----878  
Db 583 ALAQ--EARVLOGEWAHIQVGMQTQAGLLEHLKLENVLSQOLTETQHRSMKEKGRITAA 640  
QY 879 -----FHAAATVLOKWMRHWLVHVVIQRRIRHKELMRHRAATVIOAVFRGHQMK 929















/ APPLICANT: LIDGARD, GRAHAM P  
 / TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 / INTERIOR NUCLEAR MATRIX  
 / NUMBER OF SEQUENCES: 6  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 / STREET: 125 HIGH STREET  
 / CITY: BOSTON  
 / STATE: MA  
 / COUNTRY: USA  
 / ZIP: 02110  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / OPERATING SYSTEM: IBM PC compatible  
 / SOFTWARE: Patent in Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/483,924  
 / FILING DATE: 07-JUN-1995  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: PITCHER ESO, EDMUND R  
 / REGISTRATION NUMBER: 27,829  
 / REFERENCE/DOCKET NUMBER: MTP-013  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (617) 248-7000  
 / TELEFAX: (617) 248-7100  
 / INFORMATION FOR SEQ ID NO: 4:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 2101 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-483-924-4

Query Match 2.6%; Score 250; DB 2; Length 2101;  
 Best Local Similarity 18.0%; Pred. No. 1.3e-11;  
 Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

Qy	3	LWSPVLEVA-----CKETQLIDNFRKEMILKSKSNOPVKNPKFPTVG	51
Db	12	LSWNSLHVADPVAVLQDQCSFIKIIRIHGTGGQILK-----QPVSERLDFVCSF	67
Qy	52	KTQLQKSPGTAGKTMKSVSAVQOKRMSAAAAPPKQOT-----WRVTAPSR-PAWAHP	106
Db	68	LQKNRKHS-----SPECIVSA-----QKVLGSELELAKMTMLLYHSTWSSKSPDW---	116
Qy	107	PPQAPLVEKNVYKTPQEPVVI-----SPQP-----RSLKENLS	140
Db	117	-----EQFEYKIQAELAVILKFVLDHEDGLNLNEDLENFLQAPVPSTCSFPPELS	169
Qy	141	PMTGNLLDVIDNLRFTPLTETRGKGQATIPDNLAAWP-----TPTLGNVKSANDM	194
Db	170	P--PSH--QAQREIRFLELQKVASSSGNFLSGSPSPMGDIILOTPQF--MRLLKQL	223
Qy	195	RPRITPDDLDQPATNKTDFVKHSETINISLDTLDCSRIDGQPHPLNKTITIVHATHT	254
Db	224	ADERSNRDELELAENRKLTEKDAQIAMMQRID-----	259
Qy	255	RALACIHEEGPSP--PRT-----PKYSAIHDLKRDIKLVGSPILKYSE	296
Db	260	-RLALLNEKQASPLEPKEBELRDNBSLTMRLHETLKQCDLTKESQDRKINQLSE	318
Qy	297	SMKDLILLSPTKVAIQSGMFLNEMKIRSTEONRYOEQIQIKAKDLNSSSSSEASIA	356
Db	319	ENGLSFKLRFASHLQLOQALNELTEHSAKATQEWLEKQAQLE-KELSAIQDKKCLE	377
Qy	357	GQOBFLFNHSHILAQSSRFNLHEVGRKSVKSPVKNPKRSHLSHSDAPSNESLYENE	416
Db	378	EKNBIL-----QKGLSQLEE-----HLSQLQDNP-----	401
Qy	417	TVAISPPKQVDEDTTLPRSAAPANASARSSSAHAWPHAQSKFKLQATMSLMKKPATPR	476

Db	402	-----POEKEVGLGDVL-----QLETLKQEAATL	425
Qy	477	KVRDTSIQPSVKLYDSLYMQTCINPDFFAATTITDPLASTMYLDEQAVDRHQADFKW	536
Db	426	AANTQLOARVEMLETERGQGEA-----KLLAERGHFEES-----KQQ	463
Qy	537	INALVSTPADLDADANKNKIDVGLFBNVRNKLWVAPTKEBQSMNYLTKVLETLRKAHV	596
Db	464	LSSLIT--DLQSSISN-----LSQAKEE-----LEQASQAH-	492
Qy	597	ELPFSEQMRPLCSKVAVVNVKQALRISDRNLHLDVVMQRTILELLLCFNPMLWRLGLEV	656
Db	493	-----GARLTAQVASLTSSEL-----TTLNATI-----	514
Qy	657	VFGEKIQMQRNDIVGLSTFTLNRLFRNKCEEQYSKAYTL-TREYAEITKKHSLQILF	715
Db	515	-----QOQDOELAGL-----KQAKEKQAQLAQTLQOQEQASQGLRHQVQLSS	558
Qy	716	LLPFLDOAKQKRIVKNHNPCLFVKSPHKTWDIILRFSELLANIGDITRE--LRRLYGV	773
Db	559	SL-----KQK-----EQQLKEV-----ARKQEAATRODHAQQLATA	588
Qy	774	LQHRQTELDDEYAFNNL-AVDLFDGVRLTFVVEVILLRDDLTQLRVPAISLQRIENV	832
Db	589	ABEREASLRERDAALKOLEALEKEKAAL-----EILQOQLQVANEAR-----	631
Qy	833	KLALGALGEANFOLGGDIAAQDIVDGHREKT-----LSLLMQILYKFRSPFPHAAATVLIQ	887
Db	632	-----DSAQTSVTQAREKAELSRKVEELQACVETARQEQHEAQAVAE	675
Qy	888	KWRRHWHVVIQRRIRHKELMRHRAATVIAVFRGHQMKYKVLKPTERTQAAILQK	947
Db	676	-----LELQLRSEQQKATEKERVAAQ-----EKDQLOEQQLAKKESIKVTKGSLSE	720
Qy	948	FTREYLAQKQLYQSYHSIITIQWRRAQQLGRQRFVLELREAAIFLQIRWRRLPAKK	1007
Db	721	EKKR--AADALFEEQORCISELKA--ETRSVLVEQHKRRERKELEE-----ERAGRGLLEARL	771
Qy	1008	LLAA-----AETARLQRSQKQAAAAYIQ-----MOWRTYQLGRIQRHEFLQRD	1052
Db	772	LQGEAQAETEVLRRELAEMAAQHTAESCEQLVKEVAWR-----DGYEDSQEE	824
Qy	1053	L---IMVQRRMRKSWMLERKEFQOLKRAAINIQWRRAKLNMKNADYLAIRSSVL	1109
Db	825	AQYGAMFQEQMLTKBECEKARQELQEAKEKVGIESHSELQISRQ-----QNKLA	875
Qy	1110	KVQA--YRKATTQMRIDRNHYLSRKNVICLQORLRAIMQMEEQRENYLRLNASILVQKR	1168
Db	876	ELHANLARALQOQVEKEVRAQKLADDISTQEKWAATSKEVARLETIVRKAGE-----QOE	931
Qy	1169	YRMEQMIQDRNAVLRTRKCIINVQRRWRATLQMRERKNYIHLQTTTKRIQIKFRAKRE	1228
Db	932	TASRELKVEPARADR-----QPEWLEBQQGRQ-----FCSTQAAALQAMEREAQ	976
Qy	1229	MKQRAEFLQKKVTLVQKRRALLQMKREBY--LHLBEVTKLQRRFHAQKSMRPM	1286
Db	977	MGNE-----LERLAALMESQGOQOBERGQOEREVARLTQERGAQADLAL-	1022
Qy	1287	RAKYRGTOAAVSCLOMHWNRHLLKRRNSFLQLRQA--AITLQRRYARLNMKOLKS	1343
Db	1023	-----EKAARAELEMLQNAL--NEORVETATQEAALAHALTEKEGQDLAKLRLGLEA	1074
Qy	1344	YAQLKQAAITTTQRYAKKAMQKQVLYQKOREAIIKVQRYRGNLE-----MRKQI	1395
Db	1075	-AQIKLEELRQT---VKQLKEQLAKKEKHEHSSGSAQSEAAAGTEPTGPKLEALRAEV	1129
Qy	1396	-----EVQKQKQOAVIRLQKWRWSIRDMRLCKA-----GYRRIELS	1431
Db	1130	SKLEQOCQKQOQOQADSLSE--RSLEABRASAERDSALETQOQLEKQAQELGHSQSALA	1186
Qy	1432	SL-----SIQRKWRATQARQRBIFLSTIRKVRIMQAFIRATILMRQO-	1475
Db	1187	SAQRELAAPRTKQVDHSHKAEDEWKAQVARGQ-----BAERKNSLISSLEEVSLINQV	1241





Qy	1396	----	EVYQKORQAVIRLQKWRRSIRDMLCKA	-----	GYRRIRLS	14331		
		:	::	::				
Db	1130	SKLEQCCQKQEQADSLE	-----	RSLEABRASRAEDSALETLOGLEBAQELGHSOSALA	11866			
Qy	1432	SL-----	SIQKRWATVOARRQREIFLSTIRKVRMLQAFIRATILLMRQO-	1475				
			:: :	:: :				
Db	1187	SAQRELAAFRTKVQDHSKAEDEWKAQVARGQ	-----	EAEKNSLTISSLEEVSIILNRQV	1241			
Qy	1476	-RREFEMKRAAAVVI	-----	QRFRAFRCAMLKARODYQL-	IOSSVLLVORKFRANKR--	1524		
		:	::	::				
Db	1242	LEKEGESKELKRLVMAEBSKQLEBSCACCRQRPATVPLOLNAALLCRRCRASGREA	1301					
Qy	1525	----	SMKQARQEFVLRTIAVHL	-----	QOKFRGKR	-----	LMIEQCNCQPLLRCSM	1567
		:	::	::				
Db	1302	EKQBVASENLRQELTSAERABELGQELKAWQKFFQKEQALSTIQLEHTSTQALVSELL	1361					
Qy	1568	PG----	FQABARGFMARRTFQALMTPMEMDLIRQKRAAKVIQRYWRYGLYIRRRQKHQGLL	1623				
			:: :	:: :				
Db	1362	PAKHLCQQLQAEQAAAEKRHRE	-----	ELEQSKQAAGGL	----	RAELLRAQRELGELI	1410	
Qy	1624	DIRKRIALRQAEKAVNVSRCCKVQAEVRFLR	1654					
		:: :	:: :	:: :				
Db	1411	PLQKVAEQERTAQOLRAERASVAEQLSMULK	1441					

## RESULT 13

```

US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MAP
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:

```

Query Match      2.6%;    Score 248;    DB 1;    Length 2101;

Best Local Similarity 18.0%; Pred. 0.2e-11;  
Query Match 2.0%; Ccacc 2.0%; Ccacc 2.0%;  
Query Match 2.0%; Ccacc 2.0%; Ccacc 2.0%;  
Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

Qy	3	LWVSPVLEVA-----CKETQLIDNRNFRKEVMIILKSKSNQVKNPKRPEPTVG	51
Db	12	LSWVNSLHVADPVAVLQDCSIFIKIIDRIHGTEEGQOILK-----QPVSERLDFVCSF	67
Qy	52	KTLQLKSPTCAGKTMKSVVSAVQOQKRMSAAAAAPPSKOT-----WRVTAPSR-PAWAHP	106
Db	68	LQNRKHPS-----SPECVLSA-----QKVLEGSLELAKWTMLLYHSTWSSKSPRDW---	116
Qy	107	PPQAPLIVKENVYKTPQREPVYI-----SPDP-----RSLKENLIS	140
Db	117	-----EOPEYKIQIOWELAVILKFVLHDHEDGLMINEENFLQKAPVPSTCSSTFFERLS	169
Qy	141	PMTPGNLLDVIDNLRTPITETRGKQOATIPDNLAAMP-----TPTLKGNVKSICANDM	194
Db	170	P--PSH--QAKREIRFLEKQVASSGNNFLSGSPASPGWILQTPQFQ--WRRLKKQL	223
Qy	195	RPRRIITPDLOEDPATNKTFDVKHSHTINISLDTLDCRIDGQPHPLNKTTTIVHATHT	254
Db	224	ADERSNRDELELAEANRKLITKDAQIAMMOORID-----	259
Qy	255	RALACITHEEGPSP--PRT-----PTKSAIHDLKRDIKLVGSPLRKYSE	296
Db	260	-RLALINEXKQASPBPKELEBELDKNSESITWMLHETLKQCQDLKTEKSOBRKINQISE	318
Qy	297	SMKDLLSLSPTQYATIQGSPNINENMKIRSIEQNRYYOEQYQIQIKAKDLNSSSSSEASIA	356
Db	319	ENGDSLKFLREFASHLQQLDALNELTEEHSKATQEWLEKQAQLE-KELSAALQDKKCLE	377
Qy	357	GOQEFPLFNHSEIIAQSFRNHLHVGRKSVKGVKPNPKHRSHELFSAPSAPNESIYRNE	416
Db	378	EKNEIL-----QGKLSQLEE-----HLSOLONP-----	401
Qy	417	TVAISPCKQORVEDTTLPRSAAPANASARSSAHAWPHAQSKKFKLAQTWMLMKPATPR	476
Db	402	-----POEKEVIGDVL-----QLETLKQEAATL	425
Qy	477	KVRDTSQPSVKLYDSELYMQTCINDDPPAATTITDPLASTWYLDQAVDRHQADPKW	536
Db	426	AANNTQIARVEMLETERGQOEA-----KLIAERGHFEEB-----KOQ	463
Qy	537	LNALVSTPADLDADANNKIDVGKLFNEVRNKELVAPTKEEQSMNYLTKYRLETLRKAAV	596
Db	464	LSSLIT---DLQSSISN-----LSQAKEE-----LEQASQAH--	492
Qy	597	ELFFSEQMRLPCKSVAVYVNKQALRIRSDNRHLDVVMQRTILELLLCFNPLWLRLGLEV	656
Db	493	-----GARLTAQVASITSEL-----TTLNATI-----	514
Qy	657	VFEKIQMOSNRDIVGLSTILNRLPNKCEQORYSKAYTL-TEEYAEITKKHSLQKILF	715
Db	515	-----QOQDQELAGL-----KQAKKEKQAOLAOITLQOQOQASQGLRHQVQOLSS	558
Qy	716	LLPFLDAQKQIRIVKHNPCLVKKSPHKETKIDLLRFSELLANIGDITRE--LRLRGYV	773
Db	559	SL-----KQK-----EQQLKEV-----AEKQEATQDHAQQLATA	588
Qy	774	LQHRQTFIDBEDYAPNNL-AVDLRDGVRLTRVVEVILLRDDTLQRLRVPASLSQRLFNW	832
Db	589	AEEERASLURERDAALKOLEALEKEKAAL-----EILQOQLQVANEAR-----	631
Qy	833	KIALGALGEANFQJGGDIAAQDVIDGHREKT-----LSLILWOLITYFRSPKPFHAATVLQ	887
Db	632	-----DSAQTSVTOAQREKAEILSRKVEELOACVFETARQOEHAQOQVAE	675
Qy	888	KWRRHWLHVYIQRRIHKKELMRHRAATVIOAVFRGHQMKRYVKLFTKERTQAAIILQK	947
Db	676	-----LELQRSBQAKATEKERVQO-----EKDQIQEQLQALKESSLKVTGKSLEE	720
Qy	948	PTREYLAQKQIYQSYHSIIITIQWRWAQQLGRHQRFVFLERAAATFLQRIWERRERFAKK	1007
Db	721	EKRRE--AADALEBOQRCISELKA--ETRSIVFEOHKRERKELEE-----ERAGRGLEARL	771
Qy	1008	LLAA-----AETARLQRSQKQAAAASYIQ-----MOWRTYQLGRTORHEFLQRD	1052





Db 879 LGAHLGQHACAQDWQLSLBHAAQAQSAVETLQ-----APLDSLREEQRLAEALEHLQ 932  
QY 1545 OXFERGKRLMIEQRNCFQLLRCSMPGFQARGFMAKRFQALMTPEMMDLIRQKRAKVI 1604  
Db 933 QORQ-----ROQDEFQRLQADWQAWRER-QDNLDDSRDLALIG-----LSEEQATQ-- 977  
QY 1605 QRYWRGYLIRRRQK--HQGLLDIRKRIAQLRQEAQAVNSVRCKVQBAVRFLRGRFIASDA 1662  
Db 978 ---WRQLQRLQEBEITRQQTLEAERQAQLLOHRRQRPETTDREALEDNLRQQRERLAASEQ 1034  
QY 1663 --LAVLSQL 1669  
Db 1035 AYLETYSOL 1043

Search completed: September 13, 2004, 17:51:32  
Job time : 49 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:43:10 ; Search time 192 Seconds  
(without alignments)  
3108.351 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVSPVLEVACKETQLI.....FISSVAFDTLCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pap.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pap.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pap.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	672.5	7.1	898	US-10-108-260A-3399	Sequence 3399, Ap
2	408.5	4.3	726	US-10-188-832-82	Sequence 82, Appl
3	348	3.7	1022	US-10-437-963-156007	Sequence 156007,
4	300.5	3.2	1855	US-10-177-293-315	Sequence 315, App
5	291	3.1	1742	US-10-012-697-1548	Sequence 1548, Ap
6	284.5	3.0	1805	US-09-820-843A-73	Sequence 73, Appl
7	284.5	3.0	1805	US-10-282-122A-63513	Sequence 63513, A
8	277	2.9	1818	US-10-282-122A-64245	Sequence 64245, A
9	270	2.8	2383	US-10-082-830-260	Sequence 260, App
10	260	2.7	1979	US-10-205-823-419	Sequence 419, App
11	254.5	2.7	1837	US-10-369-493-22734	Sequence 22734, A
12	250.5	2.6	1151	US-10-408-765A-994	Sequence 994, App
13	250.5	2.6	1416	US-10-295-027-446	Sequence 446, App
14	244.5	2.6	1401	US-10-408-765A-2125	Sequence 2125, Ap
15	238.5	2.5	1893	US-10-408-765A-1696	Sequence 1696, Ap

16	238.5	2.5	3674	15	US-10-291-265-454	Sequence 454, App
17	238	2.5	1507	16	US-10-437-963-185157	Sequence 185157,
18	235.5	2.5	1750	14	US-10-128-714-8556	Sequence 8556, Ap
19	234.5	2.5	3899	14	US-10-171-311-4	Sequence 4, Appli
20	234.5	2.5	3917	14	US-10-171-311-8	Sequence 8, Appli
21	233.5	2.5	1965	15	US-10-369-493-3279	Sequence 3279, Ap
22	231.5	2.4	2295	16	US-10-437-963-135452	Sequence 135452,
23	230.5	2.4	1211	12	US-10-282-122A-66630	Sequence 66630, A
24	230.5	2.4	3907	14	US-10-171-311-2	Sequence 2, Appli
25	230.5	2.4	3925	14	US-10-171-311-6	Sequence 6, Appli
26	229	2.4	1601	16	US-10-437-963-185154	Sequence 185154,
27	228.5	2.4	1327	16	US-10-408-765A-1215	Sequence 1215, Ap
28	228.5	2.4	1679	15	US-10-369-493-22080	Sequence 22080, A
29	227	2.4	1707	14	US-10-128-714-3556	Sequence 3556, Ap
30	226.5	2.4	3911	15	US-10-370-685-100	Sequence 100, App
31	226.5	2.4	3911	16	US-10-408-765A-1839	Sequence 1839, Ap
32	226	2.4	3336	16	US-10-408-765A-2453	Sequence 2453, Ap
33	225.5	2.4	3225	16	US-10-408-765A-254	Sequence 254, App
34	225.5	2.4	5171	16	US-10-408-765A-2687	Sequence 2687, Ap
35	225	2.4	2052	16	US-10-437-963-137285	Sequence 137285,
36	225	2.4	3562	15	US-10-341-434-109	Sequence 109, App
37	224	2.4	1294	12	US-10-282-122A-61292	Sequence 61292, A
38	224	2.4	2020	15	US-10-369-493-5128	Sequence 5128, Ap
39	224	2.4	2020	15	US-10-369-493-5129	Sequence 5129, Ap
40	223.5	2.3	1827	15	US-10-369-493-5368	Sequence 5368, Ap
41	223.5	2.3	2310	9	US-09-874-923-120	Sequence 120, App
42	223.5	2.3	2310	9	US-09-991-496-120	Sequence 120, App
43	223.5	2.3	2310	10	US-09-820-843A-114	Sequence 114, App
44	222	2.3	1384	16	US-10-473-576-22	Sequence 22, Appl
45	221.5	2.3	2033	16	US-10-408-765A-419	Sequence 419, App

ALIGNMENTS

RESULT 1

US-10-108-260A-3399  
; Sequence 3399, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1el full length cdna  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3399  
; LENGTH: 898  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-3399

Query Match	7.1%;	Score 672.5;	DB 15;	Length 898;
Best Local Similarity	25.2%;	Pred. No. 2.3e-44;		
Matches	236;	Conservative 182;	Mismatches 309;	Indels 209; Gaps 30;
QY	850	IAAQIVDGH-----REKLSILWOLYYKFRSP-----KFHAAATVLQKWRHML	895	
Db	43	IKLOAFVGVYVRKQMLQKRAVISL--QSYFRMKARQYILKMYKAIIVIQNY--HAY	98	
QY	896	HVVIQRRTHKELMRHRAATVIOAVFRGHQMKVKVFKTERTOAAILQKFTRYLAQ	955	
Db	99	KQVNVQR---KNFLQVKKAAATCLOAAYRGYKVRQIK-----QQSIAALKIQSAFRGYNKR	151	
QY	956	KOLYQSVHSIIIFIQKRWAAQGLGRHQRFFVLEBAAFIQRIWRRLFAKLLAAAEFTA	1015	
Db	152	VKIQSVLOSIIKIQWRYRAYKTLHTRTHFLTKAAVLSLQAYRGKVRKQIRREHQA	211	
QY	1016	-----BLORSQKQ-----QAAASYIQMQRVYQLGRIGRQHEFLQRDLIMFVQ	-----	1058
Db	212	LKIQSAFRMAKQKQFRLEKTAALVIQQNFRAWTAGRCQMEYIELRHAVALQSWKKG	271	





Db 664 ERI-----LYKQKNSSISIPPIPTVTRIVSRKPDVWLRDMMBEITNPLQAIOM 717  
 QY 1852 ILCKLOI 1858  
 Db 718 VMDTLGI 724

RESULT 3  
 US-10-437-963-156007  
 ; Sequence 156007, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yinhua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 156007  
 ; LENGTH: 1022  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_55716C.1.pap  
 US-10-437-963-156007

Query Match 3.7%; Score 348; DB 16; Length 1022;  
 Best Local Similarity 19.7%; Pred. No. 3.2e-18;  
 Matches 229; Conservative 188; Mismatches 441; Indels 302; Gaps 42;  
 QY 419 AISPCKQVETTLPRSAAPANASARSSAHAWHAQSKFKLAQTMKPKPATPKV 478  
 Db 7 AVSPPRD-LSNLTNPRAFPNPKSSASKEPFP-----SATPTTR 47  
 QY 479 RDTSIQPSVKLYDSELYMTCINPDFAATTTIDPFLASTVILDEQ-----AV 526  
 Db 48 RRGPPPP-----PRCAATATPLARLRLELDQSRARRAESGRDGL 92  
 QY 527 DRHQADFCKWLNALVSIPA-----DLDA-DLANKIDV-----GKLFNE 563  
 Db 93 RAFASSATSWLSLLLRDPSACGAPSAARVTRDAPAHGVQGRDAVDGERARGSPKR 152  
 QY 564 VNKELVAPTKEEQSMNLYTKVRL-ETLRKA-----AVELFSEQMRLPCKSVAY---- 614  
 Db 153 HRGEGDRGPGPRRTTTPAMAASURDSLRVCSLDVTRMGSHMSREACEVILVMCO 212  
 QY 615 ----VNKQALIRSDNRNLHLDVVMQRTILELLCFNPLMLRLGLEVFG-----EKIQMO 665  
 Db 213 ICKNDINGLKNKEHCPLVSDLRDRKATRIPCYNPKWLRIGLHVLGDSWIQNESQK 272  
 QY 666 SNRDVGLSTFNLRLF-----RNKCEQRYSKAYTLTEEYAEATIKHSLQKILFL 716  
 Db 273 KQKEVAFLKVLKQLFVHMVTPSALNKAPEGHRRAC-----YAAASNSILKRLFL 326  
 QY 717 LPLDQAKOKR-----IVKNPCLPVKSPKETHDKILLRPSSELLANIGDITRE 766  
 Db 327 VAALDRKATESGLPSSGIDGLGGSPFLFCRQTEIKSRQIVQESLGEVHMGEGDLIMH 386  
 QY 767 LRLGLVLOHRTFLDEFDYAFNNLAVLDGVRTRVVEVILLRDLTRQLRV--PALS 824  
 Db 387 INIMGKLYNQALASEYDFTVGNLFQDGLICRIIQ--LLTSDASIIILKVIAPSDT 444  
 QY 825 RLQRIPNVKLALGALGEANFOL-----GGDIAAQDIVDGHREKTLSLIMQLIYKFRSPKPH 880

Db 445 YKRLHNCNTMAIQYIKQAGFPLSDADGLSISAEIDIYNGDKELLALLMNNFIYQPLVLV 504  
 QY 881 AAATVLOKWRHRLHVVIOIRIRHKELMRRAAAVIOAVFRGHOMRKVKVLFKTRTQ 940  
 Db 505 NETSVAQE-----ISRLKAPVSEQSI-----SEMKS 530  
 QY 941 AAILIQKFTERRVLA--OKOLYOSYHS-IITIQWRWAQQLGRQHRQFVELREAAIFLOR 997  
 Db 531 QTGLLYDWIQETLSDCRKEFLSCHTDMADIITYOFNNGIKVLAQ-----FLQD 580  
 QY 998 IWRRRLFACKLAAAEARLQSRQKQAAASYIQOMRTYQOLGRIQHEFLR----- 1049  
 Db 581 LPGWNLANDVLFVEKSA-----IILAFSLSHLTNVERLQKLNLDISKLD 627  
 QY 1050 QRDLMFVQRMRSKWSMLBQREKFOQLR-----AANIQORWAKLSMRK--- 1096  
 Db 628 HQSLVTEVSPRRSR-GTTDMKCHFPQTEETDGSRTREWAATVIQTOARRLNAMSKYCK 686  
 QY 1097 -----CNADYLALRSSLVKQAVRKATIQMIDRNHYVSLRKNVICLQORLAIM--K 1147  
 Db 687 LKNATQPCNKGHDPASSSPKSIADSSC-----IDSATKLVCEDDVDCSSNCQVLFYHD 741  
 QY 1148 MREORENYLRLRNASTL--VQKRYMRQOMIQDRNAYLRTRKCIINVQRWRATLQMRRE 1205  
 Db 742 PVSTKVDFLCRKAMAARKIQFAYRRFAHRIRSRIS-----AAIKIQSHWRC----- 788  
 QY 1206 RKNYLHLQTTTIRIQIKFRAKEMKQRAEFQOLKVTIVQKRRALLQMKRQREYLH 1265  
 Db 789 -----FSVRIFRKQI--QNTITTIQAVARLLVTGIWKRKSVIVIAHVRGWI- 832  
 QY 1266 LREVTIKLQRRFHAQSMRFMAKYRGTOAAVSCIQMHWNLHLLRKRNSFLQLRQAAI 1325  
 Db 833 ARQTAVRNKR-----ITTIQSYVKAYLLLRKRSKQ-----EITDDIM 869  
 QY 1326 TLQRYRRLANMIKOLKSYAOLKQAAITITQRYR-----AKKAMQKQVVLVYQ 1372  
 Db 870 CLINRLIAAVSQ-----RSISTIRQICATLSTATEHSEKCCQITIVNAGAVEILLKQINLEN 925  
 QY 1373 K--OREALIKVORRYRGNLEMRKQIEVVQKORQAV-IRLOKWRWSIRD-----MR 1419  
 Db 926 RGVDPQVQLKQVLFTRNIAFRNLQPVLANTPQAVEIVFQBLLRKSTEGFFVACDILKR 985  
 QY 1420 LCKA--GYRRIRLSSLSIOR 1437  
 Db 986 LCSEBEGHKLARALKRHR 1005

RESULT 4  
 US-10-177-293-315  
 ; Sequence 315, Application US/10177293  
 ; Publication No. US20030124128A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Gannavarpu, Manjula  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Mertens, Maureen  
 ; APPLICANT: Myer, Vic  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Menahan, John  
 ; APPLICANT: Meyers, Rachel E.  
 ; APPLICANT: Bast Jr., Robert C.  
 ; APPLICANT: Hortobagyi, Gabriel N.  
 ; APPLICANT: Pusztai, Lajos  
 ; APPLICANT: Meric, Funda  
 ; APPLICANT: Sahin, Aysegul  
 ; APPLICANT: Mills, Gordon B.  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
 ; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-038



```

1 / APPLICANT: Leshkowitz, Dena
2 / APPLICANT: Kita, David
3 / APPLICANT: Garcia, Veronica
4 / APPLICANT: Jones, Lee William
5 / APPLICANT: Stache-Crain, Birgit
6 / TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
7 / TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE
8 / FILE REFERENCE: 2300-16252
9 / CURRENT APPLICATION NUMBER: US/10/012,697
10 / CURRENT FILING DATE: 2003-01-21
11 / PRIOR APPLICATION NUMBER: 60/254,648
12 / PRIOR FILING DATE: 2000-12-07
13 / PRIOR APPLICATION NUMBER: 60/275,668
14 / PRIOR FILING DATE: 2001-03-13
15 / NUMBER OF SEQ ID NOS: 1568
16 / SOFTWARE: FastSeq for Windows Version 4.0
17 / SEQ ID NO 1548
18 / LENGTH: 1742
19 / TYPE: PRT
20 / ORGANISM: Homo sapiens
21 / FEATURE:
22 / NAME/KEY: VARIANT
23 / LOCATION: 1015
24 / OTHER INFORMATION: Xaa = Any Amino Acid
25 / FEATURE:
26 / NAME/KEY: VARIANT
27 / LOCATION: 1015
28 / OTHER INFORMATION: Xaa = Any Amino Acid
29 / FEATURE:
30 / NAME/KEY: VARIANT
31 / LOCATION: 1015
32 / OTHER INFORMATION: Xaa = Any Amino Acid
33 / FEATURE:
34 / NAME/KEY: VARIANT
35 / LOCATION: 1015
36 / OTHER INFORMATION: Xaa = Any Amino Acid
37 / FEATURE:
38 / NAME/KEY: VARIANT
39 / LOCATION: 1015
40 / OTHER INFORMATION: Xaa = Any Amino Acid
41 / FEATURE:
42 / NAME/KEY: VARIANT
43 / LOCATION: 1015
44 / OTHER INFORMATION: Xaa = Any Amino Acid
45 / US-10-012-697-1548

```

Query Match	3.18;	Score 291;	DB 15;	Length 1742;
Best Local Similarity	17.88;	Fred. No. 2.9e-13;		
Matches 292;	Conservative 275;	Mismatches 513;	Indels 560;	Gaps 77;
415	NETVAISPPKKQORVEDTTLPSRAAPANASARSSSAHAWPHAQSKFKLAQOTMSLKKPAT	474		
DB	415	NETVAISPPKKQORVEDTTLPSRAAPANASARSSSAHAWPHAQSKFKLAQOTMSLKKPAT	474	
382	SEIV-----VKPMTRQAVNARDALAKKIYAHLFDFIVE--RINQALQPSGKHQT	429		
475	PRKYRDTSIOPSVKLYDSELYVMQTCIN-----	501		
430	FIGVLDIY---GPEFTDVNSFEQFCINYANEKLOQOFNMHVFKLEQBEYVMKEDIPTWLLID	486		
502	-PPFPAATTITDPELASTWYLDQAVDRHQADFKKWLNALVSIADLDDADLNKKIDVVKGL	560		
487	FYDQNPVIDLIEAKMGILELDEBCLPHGTD-ENWLQKL-----YNNVFNRNPL	535		
561	FNEVR--NKELVV---APTKEEQSMNYLTKYR-----LETLEKAAVEL---PFSEQMR	605		
536	FEKPRMNTSFGVIQHFADKVEYKCEGFELEKNRDTVYDMLVEILLRASKFHLCANFFQENPT	595		
606	LPCKSAVYVYNNQALRIERSDENLHLDVVMQRTILELLCNFPLWRLGLEVVFEKLTQM	665		
596	PP-----SP-----FGSMITVVK	607		
666	SNRRDVLGLSTFILNRLFNKCEQRYSKAYTL-----TEEYAEITKCHSLQXILFLP	718		
608	SAKQVIKEN---SKHFRITVGSKFRSSLYLLMETNLATPHYVRCIKPNDEK-----LP	658		
719	FLDAQKQKRVKVNPNCLFVVKSPKHEKYDILLRFSSELLANIGDITRELRRLGVYVLOHRQ	778		
659	F--EFDKSRIVQQLRAACGVLET-----IRISAOISYPS-----RW	690		



RESULT 7

```

US-10-282-122A-63513
/ Sequence 63513, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 63513
/ LENGTH: 1805
/ TYPE: PRT
/ ORGANISM: Mycoplasma genitalium
US-10-282-122A-63513

Query Match          3.0%; Score 284.5; DB 12; Length 1805;
Best Local Similarity 19.2%; Pred. No. 1e-12;
Matches 310; Conservative 264; Mismatches 611; Indels 433; Gaps 75;

QY 296 ESMKDLSSLPOTKY-----AIGSMENLNMKIRSTEQRYVQ-----EQOIQIK 341
DB 208 ENQKRLSLEYNTYRELVSANDELQNVENIDQOIQKHQYQTYRDELSELQERKIQLT 267
QY 342 AKDLNSSSSSEASLACQGBFLFNHSEILAQSRFNLHEVGRSKVSGSPVKNPHKRSHSL 401
DB 268 KQELVDKESALRVKIDDAFYINAR--LAE-----LDDVAK-----QL 303
QY 402 SFSD--APNESLYRNETVAISPPK-----KQVDETTLPESA-APANASARSSSAHAWPH 454
DB 304 SFQDGTITKQNAQVEDKLVANKEKDRLNTQKEAFNLRQASALIDINKLQENELFA-KH 362
QY 455 A--QSKKEKLAQMTSLMK-----KPATPRKVRDTSIQPSVKLYDSLYMOTCINPDPPAAT 508
DB 363 LEHQNEFKQKSDSLKLETEYKALQHKINEFNSATK--SEELLNQ-----ERELFEKR 417
QY 509 TTIDPFL--ASTMYLDE----QAVDRHQADFKKWLNALVSIPADLDADLN-----NKIDV 557

```

```

DB 418 REIDTLLTQASLEYEHQRESSQLLKDKQNEVKHQFNQLEVAKKELDKERNLLDQKKVDS 477
QY 558 GKLFNEVRNKLWVAPTKEQSQMNYLTKYRLETLRKAAVELPFSEQMR-----LPC 608
DB 478 EAIF---QLKEKVAQRKELEEL--YLKVKQKQDKENEL--LFPEKQLKHQADFENELEA 532
QY 609 SKVAVVVKOALRIRS-----DNLHLDVVMORTILELLLCFNPLMLRLGLEVVFGKIQ 663
DB 533 KOELFPEAKHALE-RSFIKLEDKEDLNTKAQO-----IANEFSQ 571
QY 664 MQSNRDIVGLSTFILNRLFRN-KCEQRYSKAYTLTEEYAEITIKHSLQKILFLPLDQ 722
DB 572 LKTDKSKSADFELMLQNEYENLQEKQLFOERTYFERNAAVLNRLQKRELL-----626
QY 723 AKQKRVKENPCLVFKKSPHKETKDILLR-PSSELLAN-----IGDTRELRRLRYVL 774
DB 627 -----QOKETLDQLTQSFQERLINQREHKELVASVEKQKEILGKKL 668
QY 775 Q-HROTFLDEFDYAFNNLA---VDLMDGVRLTVVVEVILLRDDLTQLRVPALSRQIRIF 830
DB 669 QDFSQTSLN-----ASKNLAEEMAIKFEKEIEATEKQLLND-----VN 708
QY 831 NVKLALGALGEANFOLGSD-----TAAQDIVDGHREKTLTSLWLQIYKFRSPKFAAAATV 885
DB 709 NAEVIOADLAQLNQLNQSRLQERSLQNAKQRIADPHNDSLKKL-----749
QY 886 LQKWRERHMLHVVIQRIIRHELM-----RRHRAATVIOAVFRGHQMRKVVYKLPKERTQA 941
DB 750 -----NEYELSLQKLELOTLLEANQKHSYQN--QAYFEG-----ELDKLNREKQ 793
QY 942 AI--ILOKETRRYLAQKOLYQSVHSITTIQWRWRAQOLGROHQRQFVELREAIQRIW 999
DB 794 AFLNLKKTQMEVDAIKQRLSKHQALNNQC-----AELDRKTH-----ELNNA--FLNMDA 843
QY 1000 RRLFAKLLAAAEATAR---LQRS-----OKQAAASYI---OMQW-----RTYQIGRI 1042
DB 844 DQKSLQDQLATVKETQKLDLERSALLEKQREFAENVAGFKRHSNKTSLQKIELTKK 903
QY 1043 QRHEFLRQDRLIMFVORMRKSWMLEQRK--EFQOLKAAINIQRWRAKLSMRKCNAD 1100
DB 904 QSEQOTKETELKIAFSDIQDYQVPELOKQDBFRQ-----IEAKQRELDKLA-EKNNOV 957
QY 1101 YLALRSSVLKVOAYRKATIQMID-RNHVYSRKKNVICIQRRLRAIMKWRQ-----1151
DB 958 KLELDRFOALQKQKQDITVQAOLELREHQLNLEQTAQFQANESLQKREQLTKKIQAF 1017
QY 1152 -----RENYLRLNASILVQKRYMRQMIQDRNAYLRTKCIINVORRWRAATLQWRE 1205
DB 1018 HVELKKRNQFLALKGKRLFAKQDQQRKD--QEIN-----WR-----FKQF 1056
QY 1206 RKNYLHLQTTTKR-----IQIKFRAKREMKQKQAEFLQKLVKVTLVVQKERRAL 1253
DB 1057 EKEYTDFDAKRELELEKIRRSLSQSNVELERKEKELATFTNLNKVQHNQINRD--1114
QY 1254 LQWRKERQBYLHLREVTIKLQRRFHAOKSMRFR-----AKYGTQAAVSCLOMHWRN 1306
DB 1115 -QNSQIROFLERKNFQFSPNEANAKAFILKRLSPASNLKIQEALAIQKLEPDKRD 1173
QY 1307 HLRLKRRNSFLQLQAAITLQRRYRARLNMILKQKSAVLAQKAAITTIQTRYR-----AKK 1362
DB 1174 EQQKELQQAATLQEQFPEKQ-----NPDIEKQ-----RQLVAIKTQCEKLSDEKK 1220
QY 1363 AMQKQVVLVQKREAIL-----KVQRRYRGNLEMRKQIE-----VYQKQRAVVI 1406
DB 1221 ALNOKLVELKNLSQTYLANKNKAESQQLOQKYTNLLDKENLERTKQDLQKHSIEFA 1280
QY 1407 BLQKWRSRIR-----DMRLCKAGYRIRLSLSLQKREKATV-----1443
DB 1281 RUTKFANDURFEKKQLLKAQRIVDKRLNKENERNLHPLSNTERK-RAVLDEQISYFE 1339
QY 1444 -QARRQREIFLSTIRKVR-----LMQAFIRATLL-----MRQQRREFFEMKRAAV 1487

```

Db 1340 KORQATDAILASHKEVKKEGELQKLLVELETRKTLNNDFAKFSRQREFFENQRLKLL 1399  
 Qy 1488 VIQRFRARC-----AMLKARQDYQLIOSSVILVQRFRANS-----MKQARQBFV 1534  
 Db 1400 ELQKTLQTNNSNFKTKAIQEIENSYKRGMEELNFKQKPEFDKNSRLIYEFRRORDEIE 1459  
 Qy 1535 QLRT-IAVHLOQKPRGRKRLMEQRCNCFQLLRCSMPGFQARAGF-----MARK 1581  
 Db 1460 RKESQVKLVKETOAKNLEAQAANKLNIEKNTIDFKEKELKAPDKVDQDIDSTNKQRK 1519  
 Qy 1582 RFQALMTPEWM---DLIRQKAA-----KVTQVRWGVILRRKHQGLLDIRKRIA 1630  
 Db 1520 ELNBLNENKLLQOSLIERERAINSKDSLNNKIETI-----KROLH-----DKEMRVL 1568  
 Qy 1631 QLRQBAKAVNSVRCKVQEAVERFAGRFIAPDALAV-----LSQDLRLSRIVPHLLMW 1682  
 Db 1569 RLVDKMLAQO---KYOTEINRLRTQTFDSEKQDKIKNPPPLFKINGNDMAFFVLYPW 1623

RESULT 8

US-10-282-122A-64245  
 ; Sequence 64245, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 64245  
 ; LENGTH: 1818  
 ; TYPE: PRT  
 ; ORGANISM: Mycoplasma pneumoniae  
 US-10-282-122A-64245

Query Match 2.9%; Score 277; DB 12; Length 1818;  
 Best Local Similarity 18.0%; Pred No. 4.1e-12;  
 Matches 311; Conservative 284; Mismatches 596; Indels 536; Gaps 75;

Qy 273 TKSAIHDLKRDIKL-----VGSPLRKYS-----SMKDLISLSPQKYAIQGSMPNLNE 321  
 Db 272 TKQELVDEBTLKVRINDADFYINSRLAELDLTSKINERDFVSKEQADVKASLANLTK 331  
 Qy 322 MKTR-SIEQNRVYQEQIQIKAKDLNSSSSSEASLA---GQEEFLNHEISLAQSSPNL 377  
 Db 332 EKERLSAEKDSF--ERLUNTALNDINRMEQENALFAKHEEQOQYEFERKQ---QESLLKL 386  
 Qy 378 HEVGRKSVKSPKVNPKHRSHELFSFSDAPSNESLYRNETVAISPPKKQKQVEDTTLPRSA 437  
 Db 387 -ETHKQL-----QKRIEKFIESEAKSEALLIQERELL---EKREIDILLTQAS 433  
 Qy 438 APANASARSSSAHAWPH-----AQSKTKFLAQTMSLMKKPA 473  
 Db 434 LEYEQORRTNQVLKEKHQVQHPQNLVHAKKLDQKRHYLABQKRIDEEQIFKLKEKIA 493  
 Qy 474 TPKVVRTSTQPSVKLYDSELYMOTCINPDPFAATTIDPPLASTMWLDEQAVDRHQAD- 532  
 Db 494 TERELE-----KLY-----LVKKQKQDQKENDL 517  
 Qy 533 --FKWLNALVSI PADLDADLNKKIDVGKLFNEVRNKKELVVAPTKEPQSMNLYTKYLET 590  
 Db 518 LIFEKQLR-----QYQADFENEI-----EEKQNELPASQKSLOKSFQTKNKEAEL 563  
 Qy 591 LRKAAVELFFSEQWRLPCSKVAVYVVKQALRISDRNLHLDVVMQRTILELLCFNPLWL 650  
 Db 564 NQKA-----QKIA-----EDWAHLKQKHHAD-----L 587  
 Qy 651 RLGLEVVFGKIQMSNRDIVGLSTF-----ILNRLFRNKCE-----QRYSKAYTL 697  
 Db 588 EIFLEGEFNH-LQOEKHKLLLEARTQFQDNVRSLLSARFQKQAEVLVQKOSLEQUTAPNK 646  
 Qy 698 TEYEAETIKKHSLOKILFLPFLDQAKQKRIKVNPCFLFVKKSPHETK-----DILLRFS 753  
 Db 647 EQEAVERDWDRLANL-----EKQKMLGDKVHQFDENSLNISKLAERELAIKF 697  
 Qy 754 SELLA-----NIGDITRELRKLVGVQ-----HRTFIDDEFYAFNNLAVD 794  
 Db 698 EKELEAAQKQLSLDNNNNAGLKQLDKLSLKTERTLELEASKERILDFYDESSRIA-D 756  
 Qy 795 LRQGVRLTRVVEVILLRDDLTQLRVPALISLQRIFNVLKALGALGANFOLGGDIAQD 854  
 Db 757 YESDLQ-ARLAEVKTEKN---QOETAASERE-----LKVALEKLNQA-----796  
 Qy 855 IVDGHRKRTLSLWLQIYKFRSPKFAAAATVLQKWMRRHRLHVVIQRRIRKELMRHRA 914  
 Db 797 -----KKAFLQIRKQOLLEIASVK-----QQLAQKANLLKNQQA 830  
 Qy 915 ATVIQAVFRGHQMKYKVLKFTERTQAAIILQKFTRRYLAOKLYQSVHSIITQWRWA 974  
 Db 831 -----ELDK-----QTELEAAFLQDQTDK-----KELEKALHSV-----860  
 Qy 975 QQLGRQHRQREVELREAAIFLQRIWRRRLFAKLLAAAEATARLQORSQKQAAAAYIOMOW 1034  
 Db 861 -----KSKQELLE--RERSFLQK-----QREFAHVAGKRVQVHF 894  
 Qy 1035 RTYQLGRI-----QRHEFLRQDRL-IMFYQRRMRSKWSMLQEKPEFQQLKRAAINIQ 1086  
 Db 895 KTTQMRLSHNKQKQSQEQIKETELKATFADLKKDYQLFELQKNQBFQ-----IEQ 947  
 Qy 1087 RWRKLSMRKCNADYLALRSSVLKQAVRKA-----TIQWRID-----RNHYLSLKN 1134  
 Db 948 KHK-----ELELLAQKQAEKQLEQKATALASQDQTVQAKLDLARQOHELELRQ 999  
 Qy 1135 VICIQQRLRAIMKMQREORENVRLRNASILVQKEYR---MEQOMIORDNAVLRTRKCLIN 1191  
 Db 1000 --AFNQASLSLNKOREQLTNQVKVHGL--KKHEKLTIKDRLIAEKEKDOHKDAEIN 1055  
 Qy 1192 VQREWRATLQMRERKNYHLQTTTKR-IQIKFRAKREMKQRA-----EFTQL 1239  
 Db 1056 -QR-----PKQFENEYADFDQAKKRELQELNQRNLEQSNASLLKRNQTLTDLDFALL 1107  
 Qy 1240 KKVTLVQKRRRALLQMKERQEVYHLREVTIKQRRFHAQKSMFMEAKYGTQ-----1294



Db 1108 KVVQHTQINR---VQMTQIKFLEKKNFQKASDEALQKALLIKRLRSFASKLQQR 1164  
QY 1295 --AAVSCIQMWRNHLKRRNSFIQURQAATITQRRYRARNIMIKQKUSYAO----- 1346  
Db 1165 EALAIQKLEFDRBQKQSEINNAKIQLEFQKLEKQNFDEAKQKQIIEFQDQCORLDVEK 1224  
QY 1347 --LKQAATITQIRYAKKAMQKVVLVYKQREALIKVQRYRGNLEMRKQIE----- 1396  
Db 1225 RLLKQKLVOLKWLKSYLYTKNRADLSQOO-----LQHKYANLJELKEKQIQTAKRALDK 1278  
QY 1397 ----VYQQRQAVIRLQKWRISIRDMRLCKAGYRIRLSSLSIQR-----KWRATVQARRQ 1448  
Db 1279 KIRAIYKMAQFVSEL-----RQEKQLLSAQKQVDDKSRLLQGNQRH 1321  
QY 1449 REIFLSTIRTKVRLMQAFIRATLLMRQORREFPEMKRPAAVIQRFRFARCAMLKARQDYOL 1508  
Db 1322 LQNLSETKKRQS-----LEHDINKFDQRRKEAV----- 1351  
QY 1509 IOSSVILVQKFRANRSMKQARQEFVQLTIAVHLQOKFRGKRLMLEORNCFTOLLRCMSP 1568  
Db 1352 ----SSIL-----NSHKKLQKQEGELQIL-----OKLSLKTQIEQ--FSKL----- 1389  
QY 1569 GFQARGFEMARKRFQALMTPEMDLIQKRAAKVIQRYWRGYLIRRRQKHQGLLDI--- 1625  
Db 1390 -YQOREK-----LDQRITLTKLHRELKQNEATAHKNEVLEINY 1430  
QY 1626 -RRTAQRLQBAKAVNSVCKVQEAUFURGRFIASDA--LAVLSQDLRLSRVPHLLMW 1682  
Db 1431 YKKEQLRTTTEKSEFDNNKNRILFEYPRKIRNETEKEAHIKTVLETKKR-----HL--- 1483  
QY 1683 CSEFMSFTFCYGMQAIRSEVDKQIL-----ERCSRIILNARYNSTVTNTVFOE 1731  
Db 1484 -----VETRAVKHLQKQSIISKGQELKEIKERVSRDISITNKQRELSLHLQ 1532  
QY 1732 GGLV--TIAQMLLRCKDSIEFTNLCTLIWVFAHCPKRII--DYMTNPEAIYMWRE 1787  
Db 1533 NKLLQKNLAEREREINNKDSLITQKIQAKQLSE--KEARILKLEKMRVAVEQQYQA-E 1589  
QY 1788 TKKLVARKEKMQNARKP--PMTSGRYKSQKINFTPCSLSPLEDPF 1832  
Db 1590 ITRLKTRNADLEKNDKHLFPLF-----KINGNDMNPYPYPWF 1629

## RESULT 9

US-10-082-830-260  
; Sequence 260, Application US/10082830  
; Publication No. US2003007604A1  
; GENERAL INFORMATION:  
; APPLICANT: Sun, Yongming  
; APPLICANT: Recipon, Herve  
; APPLICANT: Salceda, Susana  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Turner, Leah  
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific  
; FILE REFERENCE: DEX-0249  
; CURRENT APPLICATION NUMBER: US/10/082,830  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,802  
; PRIOR FILING DATE: 2000-10-27  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 260  
; LENGTH: 2383  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-082-830-260

## Query Match

Best Local Similarity 2.8%; Score 270; DB 14; Length 2383;  
Matches 337; Conservative 306; Mismatches 605; Indels 512; Gaps 79;

QY 223 NISLDTDCSRIDGPHPTLNKTTTIVHATHTRALACIHERBGPSPRPTTKSAIHLDKR 282  
Db 343 SLELDSSIFSQDYQ---DADKALILVRSVLT---RQAVQDLRQ 382  
QY 283 DI-----KLVGSPLRKYSEMK-DLSLLSPQTKYAIQSGMNPNNEMKI 324  
Db 383 QLACQCEAVNLQOQHDQWEEBEGKALRQLKLTGERDTLAGQT-VDLQGEVDSLSKER- 440  
QY 325 RSEIQNRYYQEOQIQI-----KADLN-----SSSSSEASLAGOQEFIFNHSEILAQSS 373  
Db 441 ELLQKARBELRQOELVLEQEAWRRLRRVNVVELOQDSGAQOQKEBQOEFELH-----LAVRE 495  
QY 374 RNLHEV-----GRKSVKGVKPNPKRRSH-----ELSFSDA 406  
Db 496 RRELQMLNGLEAKQSELSELITIREALESSELGELLRQEQTEVTAALARAQSI AEL 555  
QY 407 PSNESLYR-----NETVAI-----SPPKKQQRVEDTTLPR 435  
Db 556 SSENTLKTVEADLRAAAVKLSALNEALALDKVGLNQQLLEENQSVCSMEAAEQAR 615  
QY 436 SAAPAN-ASARSSAHAW---PHAQSKFKLAQWLSMKKPKATPKVVRTSIQPSVKLYD 491  
Db 616 NALQVDLAEAEKRRREALMEKNTHLEAQLOKAEAGAELO--ADLRDIOBEKEIOKKUSE 673  
QY 492 SELYMQTCINPDPPFAATTTIDPF-----LASTMYLDE-----QAVDR 528  
Db 674 SRHQE-----AATTQLEQLHQEAKQOEVLARAVQKEALVREKAALVRLQAVR 725  
QY 529 HOADFKKMINALVSPADLDADL-----NNKIDVGLFNEVRNKELVVAPTKESQSMY 582  
Db 726 DRQDLAEQLOGLSSAKELLESSLFAEQONSIVETVTKGLEVQIQTVTQKEVIOGEVRC 785  
QY 583 LTKYLETLRK-----AAVELFFSEQ--WRLPCSKVA--VYVYKQALRISDRNLH 629  
Db 786 L-KLELDTERSQAQERDAAARQLAQAEQKGTALAEQQAHEKEVYNQREKWEKERSH 844  
QY 630 LDWV-----MORTILELILCFNPLWRLGLEVVFGEKIQMQSNRDI VGLSTFILNRLFR 683  
Db 845 QQELAKALESLEKWE-----LEMRLKEOQTEMEAIOAQR----- 880  
QY 684 NKCEBQRYSKAYTLTEEYAEITKHSLOKILFILPFLDOAQ-----KRIVKHNPCLFVK 738  
Db 881 ---EBERTQAESALCQMLETEK-----ERVSLLTILQTKELADASQQLERLQRDMKVQ 933  
QY 739 KSPHKETKDIILRFSSELLANIGDITRELRRLGYVLQHRQTFLDEDFDYAFNNLAVDLRQ 798  
Db 934 KLKEQETTGIQTLOEQAQREKLAARQHRDDIALAQEBSL-----LQDK 980  
QY 799 VRLTRVVEVI-----LLRDDLTRQLR-VPAISRLOQ-----IFNVK 833  
Db 981 MDLQKQVEDLKSQVAQDDSQRLVQEQVEKLERETQYNNRIQKELEREKASLTLSLMEKE 1040  
QY 834 LALGALGEANFOLGGDIAA--QDIVDGHRE-KTSLMLMOLI---YKFRSPKTHA-AATVL 886  
Db 1041 QRLVLQEADSRIRQBELSALRQDMQEAQGEQKELSAQELLRQEVKKEADFLAGEAQL 1100  
QY 887 QKWRHRHMLVVIIQRRIRHKELMRHRANTVIOAVPRGHQMKRYKVLKTE-RTQAAIIL 945  
Db 1101 BELEASH-----ITEQQLRASLWAQAKAQL-----QLRLRSTESQLEALAAE 1144  
QY 946 QKFTRRYLAQKOLYQSYHSIITIQWRRAQQLGRQHRQRFVELREAAIFLQRIW----- 999  
Db 1145 QQFGNQQAQQAQASLYSAL-----QQALGSVCESR-PELSGGSDSAPSVWGLEPQ 1195  
QY 1000 ---RRLFAKKLLAAAFARLORSQKQAAASYI---QMWNTYQLGRITORHEFLQRDL 1053  
Db 1196 NGARSIFKRGPLLTALISA-----EAVASALHKLHQLDKWKTQ-----TRDV 1236  
QY 1054 INFVQRMESKWSMLEQPK-----PFQOLKRAINIQQ---RWRAKLSMKCNADYIALR 1105  
Db 1237 LRQDVQKLEFLRTDTAEKSOVHTLEQDLQRLSQNEEKSKEGK--QNSLESELMELH 1294  
QY 1106 SSVLKVQA-YRKATIQ---MRIDRNHYSLRNKVICLOQLR-AIMKMEQ----- 1151

```
Db 1295 ETWASLQSLRRABLRQWMAQGEHLLQAQENLTAQVHLLQAQAVVBARAQAASAGILEE 1354
Qy 1152 -----RENYLRLRNASILVQRYRMRQOMIQDR-----NAVLRTRKCIINVQR 1194
Db 1355 DLRTARSAKLUKNEV-----ESERARALQOQGBELKVAQKALQENLALITQ----- 1403
Qy 1195 RWRATLQMRERKRYLHLQTTTKRIQIKFRAKRMKQARAF--LQLKKVTLVVQKRRRA 1252
Db 1404 -----TLAERBE-----EVELTGLGQI-ELEKQREMQAALELLSLDLKKRQEVLDLQOQB 1453
Qy 1253 LLQMRKQREYVHLREVTIKLQRFHQAQKSMRFMRKYRGTOAAV--SCLQMRHNLHL 1309
Db 1454 IOELEKCRSVLEHLPMVAQREKLTQVQREIRELEKDRQTRQNVLEHQLLELEKKQOMI 1513
Qy 1310 RKRRNSFLQROAAITIQ-----RRYRLANMILKOLKSYAQLKQAAIT--I 1354
Db 1514 -ESRQGVQDLKQGLVTLLECLALEENHHKWCQCKLILKEGQRETRVALTHLTIDL 1572
Qy 1355 QTRYRAKAMQOV-----VL-----YQKQREAIKVQR--YRGNLEMRK 1393
Db 1573 EERSQELQAQSSQIHDLESHSTVLARELQERDQEVKQREQIIEELQKHELTQDLERRD 1632
Qy 1394 QIEVYQORQAVIRLOKWRISIRDMRLCKAGYRIRLS-----SLSIQR-----KW 1439
Db 1633 QELMLQERIQVLEDR---TROTKILEEDLEQIKLSLRERGRELTQROLMOERABEG 1688
Qy 1440 RATVQARR---QREIFLSTIRKVRUWOAFI-----RATILM 1472
Db 1689 KGPSKAQGSLEHMKLILRDKKEVEQCEHHELOELKDLOLOGLHKKVGETSLLL 1748
Qy 1473 RQRRREFEMKRAAVIQRFRARCAMLKARDYQLIQSSVILVQRFRANRSMKQARQE 1532
Db 1749 SQREQEI-----VVLQOQLQ-----EAREQELKEQS--LOSQLEAQALAQORDQE 1793
Qy 1533 FVQLRTAVHLOQKFRGKRLMIBORNCPLLRCMPGFQARA-----RGFMARKRQOA 1595
Db 1794 LEALQ-----QROQAQGOGEERVKEK-----ADALOGALEQAHTLKERHGELODHKEQA 1843
Qy 1586 LMTPEMDLIRKAAKVIQRYWRGYLIRRKQHGGLDIRKTAOLRQE----- 1635
Db 1844 RLLEE--ELAVEGRVQALEVIGDLRAESREQEKALLALQOQCAQAEQHEVETRALQD 1901
Qy 1636 ----AKAVNSVRCKVQEA VR 1651
Db 1902 SWLQAQAVLKERDQLEALR 1921
```

## RESULT 10

```
US-10-205-823-419
; Sequence 419, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamathkar, Shubhangi
; APPLICANT: Wansley, Angela M.
; APPLICANT: Ghatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
```

```
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-419

Query Match 2.78; Score 260; DB 14; Length 1979;
Best Local Similarity 18.08; Pred. No. 1.le-10;
Matches 303; Conservative 287; Mismatches 624; Indels 472; Gaps 69;

Qy 291 LRKYSESMKDL-SLLSPQTKYAIQSGMPNLNEM-----KIRSTEONRYVQEQIQIKA 342
Db 258 LSDYEERIEBELNLQGGSGVIETDLSKIYEMQKTIQVLQIEKVSTK--KMQLEDKI 315
Qy 343 KDLNSSSSSRAS-----LAGQOEFL-FNHSEILAQSSRFNLHEVRKSVKSGPVKNPKHR 397
Db 316 KDINKKLSAENDRDILRRREQQLNVKRLQIMBECENKLE---CSKLQPSAVKQSDTMT 372
Qy 398 SHELSPSDAPNSLSYR-----NETVAISPPKQR--VEDTTLPRGAAPANASARS 446
Db 373 EKERILAQSAVEEVFLQALSDAENETMRLSLNODNSLAEDNLKMKRIEVLKEKS 432
Qy 447 SSAHAWPHAQSKFKLAQTMSLMKPATPRKVRDTSTQPSVKLYDSBYLMQTCINPPFA 506
Db 433 LLSQKEELQSLKLANNEYVIKSTAT---RDISLDS--ELHDLRLNLE-----A 478
Qy 507 ATTTIDPFLA--STMVLDQAVDRHQADFKKWL-----NALVSPADLADALIN- 552
Db 479 KEQBLNQSISEKETLIAEIBELDRQNOEATKHMILINDQLSKQONEGDSIISKLKQDLND 538
Qy 553 -----NKLDVGLFNEVRNKELVAVPTKEOSMN--YLTQYRL- 589
Db 539 EKKRVHOLEDDKMDITKEL-DVQKEKLI-----QSEVALNDHLTKQLEKVENLVQDLN 593
Qy 590 -----TLRKAAVELFSEQMRLPCSKVAVYVKNQAL-RIRSDRNHLHDVVMQRTILEL 641
Db 594 KSQESNVSIQENLEL-----KEHIRQNEEELSRI RNELMQSLQDSNSNFKDT 642
Qy 642 LLCENPLWRLGLHVEVFGKIQMSNREDIVGLSTFILNRLFRNKCEQR----- 690
Db 643 LKEREABVR-NLKQNLSELEQLNENLKVAFDVYKMEKLVACEDVRHQLBECLAGNN 701
Qy 691 -----YSKAVTLTE--BYAETIKKHS-----LQKIL 714
Db 702 QLSLEKNTIVETLKMEKGBIEAELCWAKKLLLEANKYETKIELSNARNLNTSALQLEH 761
Qy 715 FLAPFLDAQKRIKVNHPCLFVKKSPHKTKDILLFSSSELLANIGDITRELRGLYVL 774
Db 762 EHLIKLNQKKDMETAEELKKNIEQWTDHKETDVL---SSSL-----EEQQLTQLI 810
Qy 775 QHRQTFLDFEDFYAPNNLAVDIRDGVRLTRVVEVI----- 808
Db 811 NKKEIFIEKLKRSKSLQBELDKYSQALRKNEILRQTIIEKDRSLGSKMBENNHLEBELE 870
Qy 809 LLRDDLTRQLRVPALSRQLRIFNVKLALGALGANFOLGGDIA-AQDIVGHRKETSLL 867
Db 871 RLREEQSRTPADPKTLDSVTELASVQSQNTIKHELEBEIKHKKHQLIEDQNSKQWLL 930
Qy 868 WQL-----IYKFRSPKPHAAATVL----- 886
Db 931 QSLQEQKKEMDEFRYQHEQOMNATHQTLFLEKDBEIKSLQKTIQIKTQLHBEERQDITDN 990
Qy 887 -----QKWRBHLHVLIQRR-----IRHKELMRHRAATVIOAVFRGHQMR 928
```

Db 991 SDFQETKVSQNIENGSEKDLKSAETERLVKIKERELEIKLEKNISLTKQIDOLS 1050  
QY 929 K-----YVKLEKTERTOAAILLOKFTTRYLAQK----- 957  
Db 1051 KDEVGKLTQIIQOKOLEIQALHARISSTSHTDVVYLOQLOQAYAMREKVFVAVLNKTR 1110  
QY 958 ----LVQSVHSIITIQWRRA-----QQLGRHORFVE-----LREAAFIQFQWRRL 1003  
Db 1111 ENSHLKTEYHKMDVAEKAALIKLODENKKLSRPFSSGDMFRETIONLSRIREKD 1170  
QY 1004 FAKKLAAETARLQRSQKQAAASYIQMWRTYQLGRIQRHEF---LRQDLIMFVQR 1060  
Db 1171 IEIDAL-----SOKCQTLAVLQTSSTGNEAGVNSHQFEELQERDKLQOVVK 1220  
QY 1061 MSKSKSM-----LEORKEFQOLKRAANIQORWAKLSMRKCNADYIALRS 1107  
Db 1221 M-EWKKQVMTTVQNNQHESAOLELHOL-QAQLVDSDDNNSKLQV-----DYTGL---- 1270  
QY 1108 VLKVOAYRK-----ATIQMRIDR---NHVYSIRKNVICQORLRAIMKREORE 1153  
Db 1271 ---IQSYENETKLNFGQELAQVHISIQOLCNTKDLGLCKLIDTISPQLSSASLITPQSA 1327  
QY 1154 NYLRLNASILVQKRYMRQOM-----IQDRNAVLETRKCIINVQRRWR-----ATLQ 1201  
Db 1328 ECLASKSEVISESELLOLEELERKLSIQEKDATITLQ-----ENNRLSDSIAATSE 1382  
QY 1202 MRERKNYHLQTTKRIQIKFRKEMKQABFLO--LKKVTLVQKRRALLQMRKE 1259  
Db 1383 LERKE-----HEQTDSEIKOL-----KQODVLQKLLKEDLLIKAKSQQLSSNEN 1429  
QY 1260 -----RQEYHLREVTIKLQRRPHAKSMRFRMAK-YRGTOAAVSCLO--MEW 1304  
Db 1430 FTNKVNEELLQAVTNLKERLILLEMIDIGLKGNEKIVETRYGKETYSQALQETNMKF 1489  
QY 1305 RNHLKREKRNESFLOLQRAAITLORRYRARLNMIKOLKSYAQLKQAAITITQRYRAKAM 1364  
Db 1490 -SMLEKEFECH-SMKEKALAFELQKKE-----KBOGKTGELNQL-----LNAVSM 1535  
QY 1365 QKQVLYIQORBAI-IKVORRYRGNLEMRKQIEVYQKQOAVIRLOKWRSTDMRL-CK 1422  
Db 1536 QKTVVFOQERDQVMLAKQKMENTALQNEV---QRLRDKFRSNQOELERLNLHLESE 1592  
QY 1423 AGVRRIRLSLSFORKWRATVOARROREIFLSTIRKVLMOAFIRATLLMRQORREFEMK 1482  
Db 1593 DSYTREALAEDEAKLRKVVTVLEKIVSSNAMENASHQSVQVESLOQLN----- 1646  
QY 1483 RRAAVVIQRRFARCAMLKARQDYQIQQSVILVQKFRANKSMKQARQEFVOLRTIAVH 1542  
Db 1647 -----VSKQRDETALQLSV-----SQEQVKQYALSLANLQMWLEH 1682  
QY 1543 LQOKFRG-----KRLMIEORNCFOQLRCSMPGFOARAGFMARKFQALMTPEMD 1593  
Db 1683 FQOEEKAMYSABELEKOKOLIAEWKKAENLEGKVISLOECLDEANAAALDSASRLT-EOLD 1741  
QY 1594 LIRQKAAKVIQRYMRGVIILRRKXOGLL-DRKRIAOLROEKAKAVNSVRCKVQOAVRF 1652  
Db 1742 -VKEQOIEB-----LKRQNELAQEMLDDVQKKIMSL-----ANSSEKVDKVL--- 1783  
QY 1653 LRGRFI 1658  
Db 1784 MRNLFI 1789

## RESULT 11

US-10-369-493-22734  
; Sequence 22734, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22734  
; LENGTH: 1837  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-22734

Query Match 2.7%; Score 254.5; DB 15; Length 1837;

Best Local Similarity 18.4%; Pred. No. 2.7e-10;

Matches 338; Conservative 355; Mismatches 728; Indels 415; Gaps 85;

QY 203 DLEOPA--TNKTFVVKHSETINISLDT---LDCSRIDGQHPHTPLNKTITI---VHAHTT 254  
Db 159 DLKDALASTHQVLELOHTQOEKASLOIWTYEFELQKLTQKNSILENNNTWLSRELGQVND 218  
QY 255 RALACIHBEESPPRTPTKSAIHD--LKRD--KLVSPLRKYSES-----MKDLSL 303  
Db 219 KLLS-LHQE--ASLEKSQSSQLSDAVLEKDALQKVSLSQQFTESNLRYQNIVAELSE 275  
QY 304 LSPQTKYALQSWPNLNMKIRISIQ-NRYIOE--QQIQKAKDLNSSSS-----EASL 355  
Db 276 MKQOYEF--QVSF---EKEISSQKQISELMWEKCEDCSRLKELQNSGELEKLEAAQ 330  
QY 356 AQOQFELFNHSEILAQ--SRFNLHFVGRKSVKGSVPVQPHKRSH-----ELSES 404  
Db 331 SSFEQLESHKEAEASLKSQINFLKEVSSLE-SQLKANERLRHVDETEISDMSELKYS 389  
QY 405 DAPSN-----ESLYRNETVAISPCKQKQVDT-----TLPSAAPANAS 443  
Db 390 NLLNNSMGFKGQSSVSDLYSERLY-----YKQYEQTCQVERLQSY--NHVM 437  
QY 444 ARSSSAHAWPHAQSKFKPLAOTWSLMKKPATPRKVRDTSIQPSVKLYDSE---LYMOTC 499  
Db 438 EBANLQHPLVKQOFKRFAMQ-----REIVAMSEYQKSLCCKAKSRYEQ-- 484  
QY 500 INPDPPAAATTTIDPFLASTMYLDEQAVDRHQADFKKLNLALVSIQA-DLDDADLNN---KI 555  
Db 485 -----LETLFKDKCTENKHYQETKOLARQVQVLLHDLCEGVLGV 528  
QY 556 DYVKLFNEVRNKLVPVAPTKBEQSMNYLTKYRLETLRKAAVELFFSEQMRLPCKSVAVYV 615  
Db 529 DSKKKINSYVEKSL-----TEDETDTDQIISRLVWVFNIR-ELQQONQLLSA-----V 577  
QY 616 NKQALRIRSDRNHLTD--VVMQRTILELLLCFNPL-----WLRGLGVVFGKIQM 664  
Db 578 HELADRMKEDKPDLDGAEIOETLIKANETIDQTKMLEEVSDDQYKSLKRDFFRSILV 637  
QY 665 QSNRDIVLGSTFILN-RLFRNKCEQORYSKAVTLTEVYAEETIKKHSLOKILFLLPFLDQA 723  
Db 638 QENKLLDAPATPSKLNNTNLIQTSYORSIRLRLQLTNELES-----LKSIRN 688  
QY 724 KQKRVKHNPCILFVKKSPHKETKDILLRFSSELLANIGDITRELRRLGVYLOHROFTLE 783  
Db 689 KEKKFEEATSSIQLEKSNIQLOLTSLTSRSRLALEKLNDEKSL-----VLSERSK--DE 741  
QY 784 FDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTFRLEVPAISRLQR----- 828  
Db 742 LDESYSKLSQEQI-----ASKKIEV-----QNVSSQLSI-CNSQLEQSNHIVDNKSNLLL 791  
QY 829 -----IFNVKIALGALGEANFOLGDDIAAOD-----IVDGHREKTLISLLWLIYK 873  
Db 792 TSVKDKLKADLSNLSKSLSIQODNFHMAQIESSNQEYTAIVDSMNSRIELSNDL--R 849  
QY 874 FRSPKFAAATVLQKWRHRHLVVIQRRIRKHELMRRHRAATVIOAVFRGHQMRKYVKL 933  
Db 850 VANSKLSSECSDDVRR-----LTLQNSFD-----LREHQ-TLVLOLQSNITELKQDITL 896



```
QY 949 TRRYLAOKOLYOSYHSIIITQWRRAQOLGROHROFVELREAAIFLQRIWRRRLFAKGL 1008
Db 615 NRELDVKKKFDIN-----QEFVKIKKNEILKR----- 644
QY 1009 LAAAEIARLQSQKQQAASYIOMQWRTYQLGRIORHEFLRQDLIMFVQRRMRKSWML 1068
Db 645 -----NLENTQ-NQIKAEYIS-----LAEHE-----AKWSSL 670
QY 1069 EORKEFQOLKRAINIQQWRRAKLSMK-----CNADYLALRSSVLKVQAVRKATIOMRIDR 1125
Db 671 SQ-----SMRKVODSNAEILA-----NYRKG----- 691
QY 1126 NHYSLRNKVICLOLRALIMKREORENYLRNASILVQKRYMRQMIQDRNAYLRT 1185
Db 692 -----QEEIVLHAEIKAKKELDTIQECIKVKYAPI-----VSF 726
QY 1186 RKCINQRRWRATLOMRERNKYNLHLOTTTKRIQIKFRAKEMKORAEFLQLKKVILV 1245
Db 727 EBC-----ERKPKAT-----EKELKQQLSEQTKYSV-----SEEVKKNQENDKAKKEIFT 774
QY 1246 VOKRR-----ALLOMKEKEQYHLHEVITIKLORRFAOK-----SMRFMAKRYGTQA 1295
Db 775 LOKOLDRTKVLIEKSHEMERALSRTDELNKQLKDLQSKYTEVKNVKEKLVENAKOTSE 834
QY 1296 AVSCICQHWNRNL-----LRKRERNSPLOLRQAAITLORRYRARLNMILKOLKSYAQ-L 1347
Db 835 ILAVQNLQKQHVPLEQVEALKKSLNGTIENTLKEHLKSMORCYEKQQTIVTKHQLLENQ 894
QY 1348 KOAATITOTRYAKKAMQKVLYQKORBAIIKQORRYRGNLEMRKQIEVYQKQAVIR 1407
Db 895 KNSVPLAEHLQIKEAFEV-----GLIKA-----SLREKEESQNMKEEYSK 938
QY 1408 LQKWRESIDMELCKAGYRRITLSLSLQKRWATVQARQREIFLSITRKVRLQAFIR 1467
Db 939 LOS-----EVONTKOALKKQETREVPDLSKYAT-----KSDLETQISSLINE----- 980
QY 1468 ATLLMRQORREFEMKRAAVIQRFRACAMLKARODYQLIOSSVILVQKFRANRSMK 1527
Db 981 -----KLANLRKKEEVEEVLHAKKEISAKDEKLHFSI-----EQLK 1022
QY 1528 QARQEVQLTITIAVLOQKFRKRLMIEQR-----NCFOLLRCSPGFQ--ARARGF 1577
Db 1023 DQKERCDSLTTITELQRRIOESAKQIEBAKONKITELANDVERLKAQALNGLSOLYTSGN 1082
QY 1578 MARKRFOALMT-----PEMDLIRQKRAKVIQRYWGYLIRRRQKH--QGLDIR 1626
Db 1083 PTKRSQSLIDTLQHVKSLEQOLADADRQHEVIAI-----YRTHLLSAAQGHMEDVDQBAL 1139
QY 1627 KRIAQLRQ 1634
Db 1140 LQIQLMRQ 1147
```

## RESULT 13

```
US-10-295-027-446
; Sequence 446, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
```

```
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 446
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-295-027-446

Query Match 2.6%; Score 250.5; DB 15; Length 1416;
Best Local Similarity 18.1%; Pred. No. 3.7e-10;
Matches 262; Conservative 243; Mismatches 522; Indels 421; Gaps 58;

QY 277 IHDLRDKLVGSPURKYSESMDLSLLSPQTKYAIQSGMPNLNE--MKIRSIQNRYQ 334
Db 296 IQDLEIENEDLKERLKIQQEORIL-----LDKVNGLQLQNEEVMVADDESER--E 346
QY 335 EQOIQIKAKDNLSSSSSEASLAGQOEFLEFNHSEITLAQSSRFN-----LHEVGRKSVKGS 388
Db 347 KLSLLAAKEQHEESLRTIEALKNRKFKYFSDHLSGSHFNSRKBEDMLLKQGMVADS 406
QY 389 PVKNP-----HKR-----RSHLSFSDAPSNESLYENETVALSPPKQORVEDTILPSAAP 439
Db 407 QCTSGIIPAHQMSRMLRPLESL---PSQTSYSENEIL-----KKELEAMRTFCESAKQ 458
QY 440 ANASARSSSAHAWPHAQSKFKLAQTMSLMKKPPATPKRVDTSIQF-----SVKLYD 491
Db 459 DRLKLQNELAH-----KVAECKALALECERVKEDSDEQIKQLEDALKDVQKRYE 508
QY 492 SE---LYMQTCINPDPPFAATTITDFFLASTWYLDQAVDRHQADPKKWINALVSTPADLD 548
Db 509 SEGKVKQMOT-----HFLALKEHLTSEAAAGN-----HRLTEELKQOL 546
QY 549 ADLNNK-----IDVGKLFNEVRNKELVVAPTKESQSMNLYTKYRLETLRKAARELVFFSEQ 603
Db 547 KDLKVKYEGASAEVKGKRNQIKQNEMIVEEFKDEGK-----LIEENKRLQKLSMCEM 600
QY 604 MRLPCSKVAVYVKNQALRIRSDRNHLHDVVVMQRTILLELLCFNPLWLRGLGEVWFGKIQ 663
Db 601 EREKKGKRVTEMGQAKLSA-----KLSLSIAEKEFEN 634
QY 664 MQSNRDIVGLSTFTILNRLFRNKCEORYSKAYTTTEEYAEITIKKHSLOKILFLPFLDQA 723
Db 635 MKSS-----LSNEVNE---KAKLVEMEREHEK--SLSEIRQLKRELENV 674
QY 724 KOKRIVKHNPCLVFKKSPHKETKDILLRFSSEILANIGDITRELRRLGYVLOHRTFFLDE 783
Db 675 KAK-LAQH-----YKPEHEQVKSRLBQKSGELGKKITELTKNQTLQKEIE--KVLYDN 726
QY 784 --FDYAFNNLAVDLRDGVRLLRVVVEVI-----LLRDDLTRQLRVPALSRLOIRFNVKLAL 836
Db 727 KLLKEQAHNLTIEMKNHYVPLKVSDEMKSHDAIIDDLNRLK-----LDVTOKYTEKKLEEM 782
```







Search completed: September 13, 2004, 17:56:15  
Job time : 206 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 14, 2004, 21:48:33 ; Search time 15801 Seconds  
(without alignments)  
5104.828 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETQLI.....FISSVYAFDTILCKLQIDMP 1861

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DBV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US09914698/runat\_10092004\_150005\_28214/app\_query.fasta\_1.2055  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914698 @CNC 1 1 11133 @runat\_10092004\_150005\_28214 -NCFU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	9514	100.0	5586	3	DMU95171
2	9386.5	98.7	57009	2	AC019775
3	9386.5	98.7	215711	3	AC008205
4	9386.5	98.7	228433	3	AE003749
5	5267	55.4	3486	3	AY094825
6	1374	14.4	9846	10	AF533752
7	1360	14.3	10434	9	AF509326
8	932.5	9.8	3760	9	BC034607
9	834	8.8	2453	6	AX876627
10	834	8.8	2453	6	BD156229
11	834	8.8	2453	6	BD156229
12	755	7.9	4088	9	AK001380
13	753	7.9	235179	2	AC097078
14	744.5	7.8	62965	9	AY365047
15	737	7.7	5760	9	HSM80955
16	735.5	7.7	2474	9	AY099893
17	735.5	7.7	2477	9	AY099891
18	721.5	7.6	155892	6	AL353809
19	707.5	7.4	3887	6	AX876625
20	707.5	7.4	3887	6	BD156228
21	707.5	7.4	3887	9	AK001379
22	688	7.2	2847	9	AY099890
23	678	7.1	2811	6	AX833832
24	678	7.1	2811	9	AK095892
25	616.5	6.5	1950	10	BC022903
26	443	4.7	2336	9	BC040439
27	437	4.6	2414	6	AX876687
28	437	4.6	2414	6	BD156262
29	437	4.6	2414	9	AK001411
30	386	4.1	193262	3	AE007579
31	386	4.1	260367	3	AE003808
32	384	4.0	160827	3	AC099025
33	359.5	3.8	68727	3	AC004516
34	352	3.7	309506	10	AY36310291
35	350.5	3.7	36134	3	CEC45G3
36	347.5	3.7	6925	5	GGDILUTE
37	338	3.6	6599	5	GGP190MHR
38	334.5	3.5	71023	2	AC004426
39	324	3.4	9121	4	OCTRICA
40	324	3.4	308108	3	AE003473
41	323.5	3.4	9344	4	OATRICH
42	322	3.4	14800	6	AX334119
43	322	3.4	14800	9	HSU53204
44	318.5	3.3	144056	3	AC091222
45	318	3.3	5970	3	PFLSALG

ALIGNMENTS

RESULT 1

```

DMU95171      5586 bp      mRNA      linear      INV 09-APR-1997
LOCUS      Drosophila melanogaster microtubule associated protein (asp) mRNA,
DEFINITION      complete cds.
ACCESSION      U95171.1      GI:1930121
VERSION      Drosophila melanogaster (fruit fly)
KEYWORDS      Drosophila melanogaster
SOURCE      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
ORGANISM      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 5586)
AUTHORS      Saunders,R.D.C., Avides,M.C., Howard,T.I.A., Gonzalez,C. and
              Glover,D.M.G.
TITLE      The Drosophila gene abnormal spindle encodes a microtubule
              associated protein that associates with the polar regions of the
              mitotic spindle
JOURNAL      J. Cell Biol. (1997) In press
REFERENCE      2 (bases 1 to 5586)
AUTHORS      Saunders,R.D.C., Avides,M.C., Howard,T.I.A. and Glover,D.M.G.
TITLE      Direct Submission
JOURNAL      Submitted (26-MAR-1997) Anatomy & Physiology, University of Dundee,
              Medical Sciences Institute, Dundee DD1 4HN, UK
FEATURES             Location/Qualifiers
                     1..5586
                        /organism="Drosophila melanogaster"
                        /mol_type="mRNA"
                        /strain="Oregon R"
                        /db_xref="taxon:7227"
                        /chromosome="3"
                        /map="96A-B"
                     1..5586
                        /gene="asp"
                     1..5586
                        /gene="asp"
                        /function="required for the normal function of the mitotic
                        spindle"
                        /standard_name="abnormal spindle"
                        /codon_start=1
                        /product="microtubule associated protein"
                        /protein_id="AAB51540.1"
                        /db_xref="GI:1930122"
                        /translation="MELVSPVLVEACKETLQIDNRNFRKEVMILKSKNQPVKNP
                        RKPFVGTQLQKSFAGACKTWKSVSAVQKKRMSAAAPPSKQIWRVTPAPRPA
                        WAHPQAPLVEKNVYKTPQEPVVISQPSRLKNSPMPGPNLLVDINLRTPLT
                        ETRGKGQATIPFDNLAAWPTPLKGNVKSCANDMRPRITPDDLEQPATNKTDFVX
                        SETINISLDTLDCSRIDQPHTPLMKTTTIVHATHRALCIHBEEGSPPTPTKSA
                        ITHLDKRDIKLVGSPLRKYSESMDLSLSPQTKVAIOGSMPLNEMKIRSTEONRYVO
                        EQQIQIKAKDINSSSSEASLAGOQEFIFNHSIILAOSSRNLHEVGKRSVKGSPVKN
                        FHKRSHELSESDAPNSLSLYNETVALSPKKQVETITLPRSNAPANASARSSAH
                        AWPQAQKFKLQATMSLKKPEATPRKVRDTSIQPSVKLYDSELMFQICINPDPFAAT
                        TIDPFLASTMYLDEQVDRQADPFKWLNALVSI PADLDADLNNKIDVGKLFNEVRN
                        KELVVAPTKEQSNMYLTKYLETIRKAAVELFFSEQMRLLPCSVAVYVVKOALRIS
                        DRNLHLDVMVORTLELLCNPLWRLGLEVFGEKIQMOSNRDIVGLSTFILNRLF
                        RNKBEQRYSAVTLTEVAETIKKHSIQKILFLLPUDQAKOKRIVKNPLCFVKKS
                        PHKETKDILLSPSHELLANIGDITFEALRGVLVGHQOTFELDEPDYAFNNLAVDLRG
                        VRLTRVSVILLRDLTLQKLRVPAISRLQIFNVKLALGCEANFQJGDLIAQDIV
                        DGHEKTLISLLWQLYKPRKFAAATVLQKWRHRLHVVIOIRISHKELMRHRA
                        ATVIQAGFRGHMRKYVKLFKERTQAAIILQKFTRRYLAQKLYQSHYSITITQWW
                        RAQIQAGHQRQFVLEKRAATFLQIRWRRLFAKLLAAETARLQRSQKQAAASYI
                        QMWQETYOLGRDIEHFEURLDILMFVORMRSKWSMLEORKEFOOLKRAAINTQORW
                        RAKLSMRKNQDIALRSVLKQVAKATLOMRIDRNHYSLRKNVILCLOLRRAIM
                        KUREQENYLRUNASILVQKRYNRQMIQDRNAYLTRKCLINQVREWATLQMR
                        ERKNYHLITQIRIQLPRAKREMKQRAEFLQKLVTLVQKRRALLQMRKEQ
                        YIHLFEVITIKQRRFHAQKSMFRAKRYQAAVSCLOMWRNHLRLRRNSPLQJ
                        ROAATLQRRYLRANIMIKQASYLQKAAITIQTRYRAKAKMQKQVLYQKQBEAI
                        LKQVRYGNLERNMQIIEYQORQAVIRLQKWRISIRDMRLCKAGYRIRLSLSIQ
                        PKWATVQARQOREIIFLSTIRKVLMOAFIRATLLMQORREFEMKRAAVVIORRFR
                        ARCAMLKQARQQLIQSSVILVORFRANRSKMOARQEFVOLRTTAVHLQKFRKRL
                        MLEQRNCFQLLRCSMFQFQARFWMARKRFOALMTPMEMDLIRQRAKVIQRYWRG
                        YLIRRRQHQGLDITRKIAQURQBEAKAVNSVRCKQBAVRFLGRFRFTASDALAVISQ
                        LDRLSRTPVHLWMCSEFMFTFCYGINAQAINSEYDKQILIERCSRILINLARINSTTY

```

```

NTPEGLGLVTIAQMLLRWCDDSEIFNTLCTLIWFAHCKPKKKIKIHDMYMNPAIMY
VRETKKIVARKEKMKONAKPPMTSGRYKSQINFTPCSLPSLEPFGIIRYSPYTF
ISSVYAFDITLCKLQIDMF"

ORIGIN

Alignment Scores:
Pred. No.:      0      Length:      5586
Score:          9514.00      Matches:      1861
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels: 0
DB:              3      Gaps: 0

US-09-914-698-1 (1-1861) x DMU95171 (1-5586)

Qy      1      MetGluLeuValTTPSerProValLeuGluValAlaCysLysGluThrLeuGlnLeu120
Db      1      ATGGAGCTAGTGTGGAGCCCGCTCTTGGTGGCTGCAAGAGAGAGCTGCACCTAATA 60

Qy      21     AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40
Db      61     GACGACCCGCAACTCCGAAAGGAGTGATGATCATCTCAAGTCCAAGAGCAACGACGCG 120

Qy      41     ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60
Db      121    GTCAAGAACCCGCGCAAAATTTCTTACTGTCCGCAAGACCCCTGCAGCTGAAATCCCGACA 180

Qy      61     GlyAlaGlyLysThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMet 80
Db      181    GGAGCTGGCAAGACAATGAAAGCGTGTATTCCTCTGCTGCGACAAAGAGGCGCATG 240

Qy      81     SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro 100
Db      241    TCTGCAGCAGCAGCGCCTCCCTCCCAAGCAGACATGGCGAGTGACTGTCTTCCGTCCC 300

Qy      101    AlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLysThr 120
Db      301    GCTGATGGGCGACATCCACCTCCACAGCTCCTCTTGTTCGAGAAAGATGTATACAGACT 360

Qy      121    ProGlnGluGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer 140
Db      361    CCACAAGAGAGCCCGGTATATCATATCACCAAGCAGCTTCGACATCTGCACTCTTAAG 420

Qy      141    ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThrProLeuThr 160
Db      421    CCTATGAGCCAGGAACCTACTCGACGTGATTGACATCTGCGATTCCACACTCTCACC 480

Qy      161    GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr 180
Db      481    GAAACCCCGTGGCAAGGACAAAGCTACCATTTTCCGGGCAATCTCGCAGCCTGGCCACA 540

Qy      181    ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThr 200
Db      541    CCAACACTTAAAGGGAATGTAAATCATGTGCTTAATGATATGCGCCCGCTCGATCACT 600

Qy      201    ProAspLeuLeuAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu 220
Db      601    CCAGATGATCTAGAAGATCAGCTGCCACAAACAAACGTTGATGTAAAGCATTCGAG 660

Qy      221    ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr 240
Db      661    ACTATCAATATTTCTTGGACACCTTGAGCTGCTCCAGATCGATGGACACCCGATACG 720

Qy      241    ProLeuAsnLysThrThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle 260
Db      721    CCCCTTAATAGACACCAACCATTTGTGATGCCAGCACACAGAGCTCTGGCCCTGATT 780

Qy      261    HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280
Db      781    CATGAGGAGGAGGAGGACCAAGTCCACTAGGACGCCCAAGAGGCGCCATACACGACCTG 840

Qy      281    LysArgAspIleLysLeuValGlySerProLeuArgLysTyrSerGluSerMetLysAsp 300

```

841 AAGAGGACATTAAAGTTGGTGGCTTACCCTTACGAAAGTATTCGAGTCCATGAAAGAT 900  
301 LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsn 320  
901 TTGTCATTCTATCGCCACAACTAAGTAGTGGCAATTCAGGGTCTATGCGCTAAATCAAAT 960  
321 GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGluGlnIleGlnIle 340  
961 GAAATGAAATCCGCTCGATCGAACAGATAGATATATACCAGGAGCAGCATCCAGATA 1020  
341 LysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu 360  
1021 AAGGCCAAAGACTGAATAGCTCTCTAGTAGCGAGGCTAGTTTGGCCGCCACGACGAG 1080  
361 PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluVal 380  
1081 TTTCTATTCAACACAGTAGAGATCCCTCGCTCAGTCAGTCGTTTAAATCTCCATGAAGTA 1140  
381 GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu 400  
1141 GGTCCGAAGTCCGTTGAAGGAAGTCCGGTGAAGAAATCCTCACAGCGCCGCTCATGAG 1200  
401 LeuSerPheSerAspAlaProSerAsnGlnSerLeuTyrArgAsnGluThrValAlaIle 420  
1201 TTGAGTTTTTCGGATGCACCTAGCACGAATCATTTGACCGCAATGAAACTGTAGCCATT 1260  
421 SerProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla 440  
1261 TCCCTCTCTAAAGAGCAACGGTTGAGACACTACTCTGCCAGAAAGTGCAGCGCCGCCA 1320  
441 AsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPhe 460  
1321 AATGCATCTGCAAGAGCAGTAGTGCCACGCTGCGCCACACGCCCAATCCAGAGATT 1380  
461 LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp 480  
1381 AAGCTAGCAACAAACCATGTCATCTGATGAAGAGCCCGCCACACCCAGAAAGTCAAGGAC 1440  
481 ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle 500  
1441 ACTAGATTACGCTTCCGTGAAGCTCTATGACTCGAGCTGTATATGCAGACGTGCATC 1500  
501 AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr 520  
1501 AACCCGATCCATTGTGCAGCAACTAGCAAAATGATCCATTCTGGCACTACCATGAT 1560  
521 LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeu 540  
1561 TTGGATGAACAGGCTGTGATCGTCATCAAGCTGACTTTAAAGAGTGTAAATGCCCTT 1620  
541 ValSerIleProAlaAspLeuAspAlaAspLeuAsnLysIleAspValGlyLysLeu 560  
1621 GTCTCCATACCGCTGACCTGGACCGAGATTTAAATTAACAAATAGACGTGGTAAAGCTG 1680  
561 PheAsnGluValArgAsnLysGluLeuValValAlaProThrLysGluGlnGlnSerMet 580  
1681 TTTAACGAGGTGCGCAACAAAGAGCTCGTGGTGGCTCCCCCAAGGAGGAGCAGTCTATG 1740  
581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600  
1741 AACTACCTAAACGAAATACCGCTCGAGAGCTTCGTAAGCGGCTGTGGAGCTCTCTTC 1800  
601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
1801 AGTAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATATGTCAACAAAGCAAGCTCTG 1860  
621 ArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
1861 GCATCCGTAGCATCGTAATCTTCACCTAGACGTAGTATATGCAACGACCACTACTTGAG 1920  
641 LeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValPheGlyGlu 660  
1921 CTGCTGCTTTGCTTCAATCCCTTTGGCTGCGCTTGGACTTGAAGTGTCTTTGGCGAG 1980

661 LysIleGlnMetClnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
1981 AAGATCCAGTCAGTCTAATCGAGACATTTGTAGGCTCAGACCTTTATCTCTCAATCGC 2040  
681 LeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
2041 TTGTTCCGAAATAAGTGTGAGGAGCAGAGTACGCAAGCAGATACACTCACCGAAGAG 2100  
701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720  
2101 TACCGGAGACCATTAAGAAGCACTCATTTGCAGAAATCTCTTTTGTGCTCTTCTCTC 2160  
721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
2161 GATCAAGCTTAAGCAGAAGCGCATCGTCAAGCAATCCCTGTTTGTGTTAAAAAGTCG 2220  
741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIle 760  
2221 CCACATAAAGACACCAAGGATATTCTGCTGCGCTTCTCTGCGAGCTGCTCGCAACAT 2280  
761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
2281 GGTGATATTACGCGGAACTTCGTGCGCTAGCTAGTTCCTACAGCACCGCCCAACATTT 2340  
781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
2341 TTGGAGAGTTTGATATATGCTTCAACAACTTGGCTGTGGACCTAAGAGATGCGTGAGA 2400  
801 LeuThrArgValValGluValIleLeuLeuArgAspLeuThrArgGlnLeuArgVal 820  
2401 CTAAACCGAGTCTGGAGGTAAATTTACTACCGATGATCTAACCCGCCAGTTTAAGGGT 2460  
821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
2461 CCAGCCATCTCTCGCTTCCAGCGATCTCAATGTAAGCTAGCTCTGCGGCACATTGT 2520  
841 GluAlaAsnPheGlnLeuGlyClyAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
2521 GAAGCCAACTTCCAGCTAGCGCGGACATCGCGCCCAAGACATCGTTGACGACATCGT 2580  
861 GluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
2581 GAGAGAGCGTTTCCCTGCTCTGGCAACTATTACAAATTCGCTCGCCCAAGTTTCAAT 2640  
881 AlaAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGln 900  
2641 GCGCGCGCCACGGTCTCCAGAAATGGTGGCGCGCTCACTGGCTGCACGCTGTATCCAG 2700  
901 ArgArgIleArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
2701 CGTCGATTCGCCACAAAGAGCTTATGCTCGCCACCGCGCGCTACTGTCTATTCAGGCC 2760  
921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
2761 GTGTTCCGTGGCCACCATGAGAAAGTACGTGAAGTTGTTTAAAGCGAAGCAGCTCAG 2820  
941 AlaAlaIleLeuLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960  
2821 GCGCAATAATTCGCAAAAGTTCACCGCTCGGTATTTGGCCCAAGACAGCATATATCAG 2880  
961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
2881 AGCTATCATGATATTATCACCATCCAGCGCTGGTGGCGAGGCCCAACAATCGGAGGCGAG 2940  
981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArg 1000  
2941 CACGCCAGCGGTTGTGAGCTCGAGAGCTCGGATCTTCTTCAGCGGATCTGGCGG 3000  
1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
3001 CGAGCACTCTTTGCCAAAAAACTATTGGCGCGCGCGGAAACAGCAGCACTTCAGCGATCG 3060

QY 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGly 1040  
DB 3061 CAAAAACAACAGGCAGCTGCTAGTTATTTCAATGCAATGGCAACGATATCAGCTGGGC 3120  
QY 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
DB 3121 AGAATTCAGGCACACAGAGTTCTCTGGCCAGAGGACCTCATCATGTTTGTTCAGCGCAGG 3180  
QY 1061 MetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
DB 3181 ATGCGAAGCAAGTGGAGTATGCTGGACAGCGCAAGAGATTCCAGCAACTAAAGCGGTGCA 3240  
QY 1081 AlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
DB 3241 GCTATTAATATCCAAACCTGGCGAGCGAGCTTTCAATGAGAAAGTGCACAGCTGAT 3300  
QY 1101 TyrIleAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
DB 3301 TATTTGGCACTTCGTTCCAGCGTTCTTAAAGTTCCAGGCTTTACAGAAAGCCACAATCCAG 3360  
QY 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
DB 3361 ATGAGATATAGATCGTATCTACTATTTCCCTGCGAAAAAATGTATCTGCTGCAACAG 3420  
QY 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
DB 3421 CGACTGAGGGCCATCATGAAAAATCGCGGACAGAGGGGAAATATATCTGAGGCTGCGAAT 3480  
QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
DB 3481 GCTTCAATACTAGTTCAAAACCGTACCGCATCGTCAACAAATGATCCAGATAGAAAT 3540  
QY 1181 AlaTyrIleuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
DB 3541 GCATATTTAAAGAACCCGCCAATGTATCATCAATGTCCAGAGCGCTGGAGAGCCACTCTG 3600  
QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
DB 3601 CAGATGGCTCGAGAAAGGAGAACTACCTTCATCTCCAAACGACACAAACGAATTCAA 3660  
QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
DB 3661 ATCAAGTTCCGTCGCAAGCGCGAAATGAAAAAGCAAGAGCGAGATTCTTTCAGCTGAAA 3720  
QY 1241 LysValThrIleuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
DB 3721 AAGGTTACCTTTAGTCCAGAAACGTCGCGGGCTTTGCTGCAGATGCGAAAGAGCGC 3780  
QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
DB 3781 CAGAGTACCTACATCTCGCGAGGTTACCATATAAGCTGCAGCGCAGATTCCATGCTCAA 3840  
QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
DB 3841 AAATCAATGCGGTTTCATCGAGCAAGTACCGCGCACCCAGGCTGCTGTGAGCTGCGCTG 3900  
QY 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320  
DB 3901 CAGATGCATTTGGCGTAAATCATCTCTCTAGGAAACGGGAGAGAACAGTTTCTTACAACTG 3960  
QY 1321 ArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGln 1340  
DB 3961 CGTCAAGCAGCAATTAACACTCGAGCGAGCGGTACCGAGCTCGTCTGAATATGATCAAGCAG 4020  
QY 1341 LeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAla 1360  
DB 4021 TTGAAGATTTAGCCCGCAGCTGAAACAGGCAGCTATTACATTCAAACCCGATATAGAGCC 4080  
QY 1361 LysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleLys 1380  
DB 4081 AAGAGGCAATGCAAAAGCAGGTGCTTTGTATCCAAAAGCAAGGAGCAATATCAAA 4140  
QY 1381 ValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLys 1400

DB 4141 GTCCAACGACGATACCCGCGCAATCTGGAGATGAGGAAGCAGATTGAAGTCTACCAAAA 4200  
QY 1401 GlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeu 1420  
DB 4201 CAGCGCCAGGAGTCAATCCGCTTGCAGAAATGTGGCGCAGTATACCGCAATCGCGGCTG 4260  
QY 1421 CysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArg 1440  
DB 4261 TGCAAGCGGCTACCGCAAGATTCCAGTCTCAGTTTCATTGAGCATTCACGCAAGTGGCGG 4320  
QY 1441 AlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArg 1460  
DB 4321 GCCACAGTGCAGGCTGTCGCCAACCGGAGATCTCTTAAAGCACCATCCGCAAAAGTCGA 4380  
QY 1461 LeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGlu 1480  
DB 4381 CTTATGCGGCAATTTATCAGACCACTTTTACTGATGCTCAGAGCGCAGGAAATTCGAG 4440  
QY 1481 MetLysArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeu 1500  
DB 4441 ATGAAGCGAAGGCTGCGGTAGTGATCCAGCGCGGTTTCGTGCTCGCTGTCTATGCTA 4500  
QY 1501 LysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPhe 1520  
DB 4501 AAGCGCAGACAGGATTACCAATTAATCCAATCTCTGTGATCTCTGGTGCGAGCGCAAAATC 4560  
QY 1521 ArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAla 1540  
DB 4561 CGTCTTAATCGCAGCATGAAGCAGGACCGCAGGAATTTGTCAGCTGCGTACTATTGCA 4620  
QY 1541 ValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPhe 1560  
DB 4621 GTCCATTTGCAACAAAAGTTCCGTGGCAAGCGTCTAATGATGAGCAGCGTAATGTGTTTC 4680  
QY 1561 GlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArg 1580  
DB 4681 CAATGCTCGCGTGTTCATCGCGGCTTCAGGACAGTGCCTCGCGCTTTATGGCTCGC 4740  
QY 1581 LysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAla 1600  
DB 4741 AAAGCATTCAGGCCCTGTATGACACCCGAGATGATGACCTCATCCGCCAGAGCGCGCC 4800  
QY 1601 AlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGln 1620  
DB 4801 GCCAAGGTTATACAGCGTTACTGGCGGGCTATCTAATCCGACGACGTCAGAACGCCAG 4860  
QY 1621 GlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsn 1640  
DB 4861 GGTCTCTGGATATCCCGAAGCGTATTGCTCAGCTGCGGCAAGAGGCAAGCGGTAAAC 4920  
QY 1641 SerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSer 1660  
DB 4921 TCTGTGCGTGCAGAAATCCAGAGCGGTGGTTCCTTCGCGACGCTTATTCGATCA 4980  
QY 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu 1680  
DB 4981 GATCGTTAGCAGTGTAAAGTCAATTCGATCGTCTTTCCGCGCATCTGTGCCACACCTGCTC 5040  
QY 1681 MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg 1700  
DB 5041 ATGTGGTGTTCGGAGTTTCATGTCACGTTTTTGTGATGGCATCATGGCTCAGGCCAATTCGA 5100  
QY 1701 SerGluValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArg 1720  
DB 5101 TCAGAGTGGATAGACGCTTATAGAGCGTTCAGCGGATCATCTTAATTTGSCCGC 5160  
QY 1721 TyrAsnSerThrThrValAsnThrPheGlnGluGlyLeuValThrIleAlaGlnMet 1740  
DB 5161 TACAATAGCACACCGGTGAACACGTTCCAGAGGCGGTTTTGGTCCACCATTTGCCAGATG 5220  
QY 1741 LeuLeuArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrp 1760

```

Db      5221 TTATTGCGTGGTGGCGAAGAGCAGTGCAGATATTCAACACTTGTGTGCACCCCTCATTTGG 5280
QY      1761 ValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGlu 1780
Db      5281 GTATTGCCCACTGTCTCTAAAGAGCGAAAGATCAATTCACGACTATATGACCAACCCAGAG 5340
QY      1781 AlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGln 1800
Db      5341 GCCATTATACATGTGGCGAAGCTAAGAAAGCTGTGGCCCGCAAGGAAAGATGACAGCAA 5400
QY      1801 AsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPhe 1820
Db      5401 AATGCGACGCAAGCGCGCGCAATGACAAAGTGGACGGTATTAAGAGCCCAAAAGATAAACTTC 5460
QY      1821 ThrProCysSerLeuProSerLeuGluProAspPheGlyIleIleArgTyrSerProTyr 1840
Db      5461 ACGCCGTGTTCCCTGCGCCAGCTTGGAGCGGAGCTTCGGAATCAATCCGCTACATGCCCTAC 5520
QY      1841 ThrPheIleSerSerValTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMet 1860
Db      5521 ACGTATTATCTCGTCCGTTTACGGCTTCGATACGATTTTGTGCAAGCTGCGAGATCGACATG 5580
QY      1861 Phe 1861
Db      5581 TTT 5583

RESULT 2
AC019775
LOCUS   AC019775
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC019775
VERSION   AC019775.1 GI:6665122
KEYWORDS HTG; HTGS PHASE2.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
          1 (bases 1 to 57009)
AUTHORS  Adams,M. and Venter,J.C.
TITLE    Direct Submission
JOURNAL  Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
COMMENT  This sequence was identified as CDW:10210906 by the submitter.
          For more information on this record e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.

FEATURES
         source
         1..57009
         /organism="Drosophila melanogaster"
         /mol_type="genomic DNA"
         /db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.:      0      Length:      57009
Score:          9386.50  Matches:      1855
Percent Similarity: 96.32%  Conservative: 4
Best local Similarity: 96.11%  Mismatches: 2
Query Match:      98.66%  Indels:      69
Db:              2      Gaps:      3

US-09-914-698-1 (1-1861) x AC019775 (1-57009)
QY      1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrIleuGlnLeulle 20
Db      21900 ATGGAGCTAGTGTGGAGCGCCCTCTTACGAGTGGCGCTGCAAGAGAGCGCTGCAGCTATA 21959
QY      21 AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro 40
Db      21960 GACACCCGCACTTCGGAAGAGGAGTGATGATCATCTACTCAAGTCCAAGCAACCGCGC 22019

```

```

QY      41 Val----- 41
Db      22020 GTCAAGTGCAGTCAAAAAATGCATTTGCCACCCCTTTTACTTACATACCCTCTTTTGCAT 22079
QY      42 -----LysAsnProArgLysPheProThrValGlyLysThrLe 54
Db      22080 GTATCTCTTACCTTTCTACCTGGAAGAACCCCGCGCAATTTCTTACTGTGCGCAAGACCCCT 22139
QY      54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValValSerAlaAlaVa 74
Db      22140 GCAGCTGAAATCGCCGACAGAGCTGGCAAGACAATGATAAGAAAGCGTGTATCCCGCTGTGT 22199
QY      74 lGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94
Db      22200 GCAGCAAAAGAAAGCGGATGTCGACGAGAGCGCTCCCTCCCAAGCAGACATGGCGAGT 22259
QY      94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValG1 114
Db      22260 GACTGCTCCTTCGCTCCGCTGCATGGGCAATCCACCTCCACAGCTCCTCTTGTGCA 22319
QY      114 uLysAsnValTyrLysThrProGlnGluProValTyrIleSerProGlnProArgSe 134
Db      22320 GAAGAATGTATACAGACTCCACAAGAGAGCGGTATACATATCACCAGAGCTTCGAG 22379
QY      134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154
Db      22380 TCTTAAGGAAATCTAAGCCCTATGACGCCAGGAAACCTACTCGAGTGATTGACAATCT 22439
QY      154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174
Db      22440 GCGATTCACTCTCACCGAAACCCGTGGAAAGCAAGCTACCATTTTCCCGGACAA 22499
QY      174 nLeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMe 194
Db      22500 TCTGGCAGCTGGCCACACCAACACTTAAGAGGATGTAAATCATGTGCTAATGATAT 22559
QY      194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214
Db      22560 GCGGCGCGTCTGAATCACTCCAGATGATCTAGAAGATCAGCCTGCCACAAACAAACGTT 22619
QY      214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArg1 234
Db      22620 CGATGTAAAGCAATTCGAGACTATCAATATTTCTGTGGACCTTGAGCTGCTCCAGGAT 22679
QY      234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisTh 254
Db      22680 CGATGGACAACCGCATACGCCCTAAATAAGACAACAACCAATTGTGATGCCAGCACAC 22739
QY      254 rArgAlaLeuAlaCysIleHisGluGluGlyProSerProProArgThrProThrly 274
Db      22740 CAGAGCTCTGGCTGTATTTCATGAGAGAGAGGAGCAAGTCCACCTAGGACCCACGAA 22799
QY      274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294
Db      22800 GAGCGCCATACACGACTGAGAGAGGACATTAAGTTGTGGTTTCACCTTACGAAAGTA 22859
QY      294 rSerGlnSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnG1 314
Db      22860 TTCCGAGTCCATGAAGATTGTCACTTCTATCGCCACAACAACTAAGTATGCCATTCAAGG 22919
QY      314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyG1 334
Db      22920 GTCTATGCTTAATCTAAATGAAATGAAATCCGCTCGATCGAACAAGATAGATATTACCA 22979
QY      334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaSe 354
Db      22980 GGAGCAGAGATCCAGATTAAGGCCCAAGACTTGATAGTCTCTTAGTAGGAGCTAG 23039
QY      354 rLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374
Db      23040 TTTGGCCGGCAGCAGGAGTTTCTATTCAACCAAGTGCAGATCCTCGCTCAGTCCAGTCG 23099

```

QY	374	gpheAsnLeuHisgluValGlyArgLysSerValLysGlySerProValLysAsnProHi	394		24180	TTTGTGTTGTTAAAGATCGCCACATAAAGAGACAAGGATATTCTGCTGCGCTTCTCGTC	24239
Db	23100	TTTTTAATCTCATGAAGTAGTCCGAAGTGGTGAAGGAGAGTCCGGTGAAGATCTCTCA	23159				
QY	394	sLysArgSerHisgluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyAr	414		754	xGluLeuLeuAlaasnileGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValle	774
Db	23160	CAAGCGCGCTCTCATGAGTTGAGTTTTCGGATGACCTAGCAACAGAAATCATTTGTACCG	23219		24240	GGAGCTGCTCGCAACATTTGGTATATACGGGGAATCTCGTCCCTGGGTACGTTCT	24299
QY	414	gAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr	434		774	uGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValas	794
Db	23220	CAATGAACCTGTAGCCATTTCCCTCTCTAAAGCAACGGGTGAGACACTACTCTGCC	23279		24300	ACAGCACCGCAACATTTTGGACGAGTTTCGATATATGCTTCAACAACTTGGCTGGA	24359
QY	434	oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrrProHi	454		794	pLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAspLe	814
Db	23280	CAGAAGTGCAGCGCGCGCAATATGATCTGCAGAGAGAGTAGTCCCAOGCTGGCCACA	23339		24360	CCTAAGAGATGGGTGAGACTAACCAGATCATGGAGTTAATTTTACTACCGGATGATCT	24419
QY	454	sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh	474		814	uThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLe	834
Db	23340	CGCCCAATCCAAGAGTTTAAGCTAGCACAAACCATGTCTGATGAAGAGCGCGCAC	23399		24420	AACCGCCAGTTAAGGTGCAGCCATCTCTCGCTTCAGGGATCTTCAATGTAAAGCT	24479
QY	474	rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyArgPheGluLe	494		834	uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAs	854
Db	23400	ACCAAGAAAGTCAAGGACACTAGCATTCAGCCTTCGGTCAAGCTCTATAGCTCGAGCT	23459		24480	AGCTCTGGCGCACTTGGTGAAGCAACTTCAGCTAGGCGCGACATCGCGCGCCAGA	24539
QY	494	uTyMetGlnThrCysIleAsnProAspPheAlaAlaThrThrIleAspProPh	514		854	pIleValAspGlyHisargGluLysThrLeuSerLeuLeuTrpGlnLeuIleTyLysPh	874
Db	23460	GTATATGACAGCTGTCATCAACCGGATCCATTTGCAGCAACTACGACAAATGATCCATT	23519		24540	CATCGTTGACGGACATCGTGAGAAAGCGTTTCCCTGCTCTGGCACTTATTTACAATT	24599
QY	514	eLeuAlaSerThrMetTyLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy	534		874	eArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTr	894
Db	23520	TCGGCATCTACCATGTATTTGGATGAACAGCTGTGGATCGTCAATCAAGCTGATTTAA	23579		24600	CCGTCTGCCCAAGTTTCTGCGCGGCCACCGTCTCCAGAAATGGTGGCGCGCTCACTG	24659
QY	534	sLysTrpLeuAsnAlaIleValSerIleProAlaAspLeuAspAlaAspLeuAsnLys	554		894	pLeuHisValValleGlnArgArgIleArgHisLysGlnLeuMetArgArgHisArgAl	914
Db	23580	AAAGTGTAAATGCCCTTGCTCTCATACCCGCTGACTGCGACGAGATTTTAAATAACA	23639		24660	GCTGCACGTTGTATCCAGCGTCGCAATTCGCCACAAAGAGCTTATGCTGCCACCGGCG	24719
QY	554	sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValAlaProTh	574		914	aAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyValLysLeuPh	934
Db	23640	AATAGAGCTTGTAAGCTGTTTAAAGAGGTGGCAACAAAGAGCTCGTGGTGGCTCCAC	23699		24720	CGCTACTGTCTATTGAGCGCGTTCGCGTGGCCACAGATGAGAAAGTACGTGAAGTTGT	24779
QY	574	rLysGluGlnSerMetAsnTyLeuThrLysTyArgLeuGluThrLeuArgLysAl	594		934	eLysThrGluArgThrGlnAlaAlaIleLeuLeuGlnLysPheThrArgTyLeuAl	954
Db	23700	CAAGGAGGAGCAGTCTATGAACCTACTTAACGAAATACCGCTGGAGAGCGCTTCGTAAGGC	23759		24780	TAAAGCGAAGCGCACTCAGCGCGCAATAATTTCTGCAAAAGTTTCAACCGTATTTGGC	24839
QY	594	aAlaValGluLeuPheSerGluGlnMetArgLeuProCysSerLysValAlaValTy	614		954	aglnLysGlnLeuTyArgSerTyHisSerIleIleThrIleGlnArgTyTrpArgAl	974
Db	23760	GGCTGTGAGCTCTCTTCAGTGAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATA	23819		24840	CCAGAGCAGCTATATCAGAGCTATCAGATTTATCACCATCCAGCGCTCGTGGCGAGC	24899
QY	614	rValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMe	634		974	aglnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh	994
Db	23820	TGTCACACAGCAAGCTCTCGCATCCGTAGCGATCGTAATCTTCACTTAGAGCTAGTTAT	23879		24900	CCAAACACTGGGAAGGCGACCGCGCTTTGTGGAGCTCCGAGAGGCTGCCATCTT	24959
QY	634	tGlnArgThrIleLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLe	654		994	eLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluTh	1014
Db	23880	GCAAGCACCACCTACTGAGCTGCTGCTTTCATCCCTTTGCTGCGCTGGACT	23939		24960	TCCTTCAGCGGATCTGGCGCGACGACTCTTTGCCAAAAAATATTGGCGGCGCGGAAAC	25019
QY	654	uGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe	674		1014	rAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaIleAserTyIleGlnMetGlnTr	1034
Db	23940	GGAGTGTGCTTTGGCGAGAGATCCAGATGCGAGTCTTAATCCAGACATTTAGGCTCAG	23999		25020	AGCCAGACTTCAGCGCATCGCAAAAACAACAGCAGCTCTAGTTATATTAAATCAAATG	25079
QY	674	rThrPheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTySerLysAl	694		1034	pArgThrTyArgGlnLeuGlyArgIleGlnArgHisGlnPheLeuArgGlnArgAspLeuI	1054
Db	24000	CACCTTTATCCTCAATCGCTTGTCCGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC	24059		25080	CGCAAGCATCAGCTGGCGAGATTCAGCGACAGCAGTTCCTGCGCCAGAGGAGCTCAT	25139
QY	694	aTyThrLeuThrGluTyTyAlaGluThrIleLysLysHisSerLeuGlnLysIleLe	714		1054	eMetPheValGlnArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPh	1074
Db	24060	ATACACACTCACGAAGAGTACCGCGAGACCAATTAAGAGCACTCATTCAGAAATCTCT	24119		25140	CATGTTTGTTCAGCGCAGGATCGCAAGCAAGTGGAGTATGCTGGAGCAGCGAAGAGTT	25199
QY	714	uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCy	734		1074	eGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMe	1094
Db	24120	CTTTTGTGCTTTCTCGATCAGCTAAGCAGAGCGCATCGTCAACACATCCCTG	24179		25200	CCAGCAACTAAAGCGTGCAGCTATAAATATCAACACCTCGCAGCGAGCTTTCAAT	25259
QY	734	sIleuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSe	754		1094	tArgLysCysAsnAlaAspTyLeuAlaLeuArgSerValLeuLysValGlnAlaTy	1114



Db 25260 GAGAAAGTCACAGCGCTGATTATTTGGTCACCTCGTTCCAGCGCTTTCTTAAAGTTCAGGCTTA 25319  
QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
Db 25320 CAGAAAGCCACCAATCCAGATGAGATAGATATCGTAATCACTACTATTTCCTCGCAAAAAA 25379  
QY 1134 nValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAs 1154  
Db 25380 TGTATTCTGCTGCAACAGACGACTGAGGGCCATCATGAAATGCGCAACAGAGGAAAA 25439  
QY 1154 nTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnG 1174  
Db 25440 TTATCTCAGGCTCGAAATCTCTCAATCTAGTTCAAAAACGCTACCGCATGCGTCAACA 25499  
QY 1174 nMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnAr 1194  
Db 25500 AATGATCCAGATAGAAATGCAATATTAAAGAACCCGCAATATGATCATCAATGTCAGAG 25559  
QY 1194 gArgTTrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnTh 1214  
Db 25560 GCGCTGGAGAGCCACTCTGCAGATGCTCGAGAAAGGAGAACTACTCTTCATCTCCAAAC 25619  
QY 1214 rThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl 1234  
Db 25620 GACAAACCAACGAAATTCAAATCAAGTTCCGTGCAAGCGCGAAATGAAAAGCAAGAGC 25679  
QY 1234 aGluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLe 1254  
Db 25680 CGAGTTTCTTCACTGAAAAGAGTTACCTTTGAGTCCAGAAACGTCGGCGGGCTTTGCT 25739  
QY 1254 uGlnMetArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrIleLysLeuGl 1274  
Db 25740 GCAGATGCGAAAGAGCGCCAGGAGTACCTACATCTGCGCGAGTTACCAATAAGGTGCA 25799  
QY 1274 nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl 1294  
Db 25800 GCGCAGATTCCATGCTCAAAATCAATGCGGTTCATGCGAGCCAAAGTACCGGGCCACCCA 25859  
QY 1294 nAlaAlaValSerCysLeuGlnMetHisTTrpArgAsnHisLeuLeuArgLysArgGluAr 1314  
Db 25860 GCGTGTGTGAGTGCTCGACATGATTTGGGTAAATCATCTCTTCTTAGGAAACGGGAGAG 25919  
QY 1314 gAsnSerPheLeuGlnLeuArgGlnAlaIleThrLeuGlnArgArgTyrArgAlaAr 1334  
Db 25920 GAACAGTTCTTACAACTGCGTCAACAGCAATACACATGCGAGCGACGTTACGAGCTCG 25979  
QY 1334 gLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrI 1354  
Db 25980 TCTGAATATGATCAAGCAGTTGAAGAGTTACGCCCAAGCTGAAACAGCGAGCTATTACCAT 26039  
QY 1354 eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGl 1374  
Db 26040 TCAACCCCATATAGGCCCAAGAGCAATGCAAAAGCAGGTGGTCTTGTACCAAAAGCA 26099  
QY 1374 nArgGluAlaIleIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGl 1394  
Db 26100 AAGGGAAGCCATTATCAAGTCAACGACGATACCGCGGCATCTCGAGATGAGGAGCA 26159  
QY 1394 nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTTrpArgSe 1414  
Db 26160 GATTGAAGTCTACCAAAAAACAGCGCAGCTCATCCGCTTGCAGAAATGGTGGCGCAG 26219  
QY 1414 rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe 1434  
Db 26220 TATACCGCATGCGGCTGTGCAAGCGGGCTACCGAAGGATTCGACTCATGTTCAATTGAG 26279  
QY 1434 rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe 1454  
Db 26280 CATTCACGCAAGTGGCGGCCACAGTCGAGGCTGCTGCCCAACGGGAGATCTTCTTAAG 26339  
QY 1454 rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl 1474  
Db 26340 CACCATTCGCAAAAGTCGACTTATGAGGCAITTTATCAGAGCAACTTTACTGATCGGTCA 26399

QY 1474 nGlnArgArgGluPheGluMetLysArgArgAlaAlaValValIleGlnArgArgPheAr 1494  
Db 26400 GCAGCGCAGGGATTCAGATGAAGAGGGCTGCGGTAGTGATCATCCAGCGCGGTTTCG 26459  
QY 1494 gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValI 1514  
Db 26460 TGTCTGCTGTGCTATGCTTAAAGCGGAGACAGGATTTACCAATTAATCCAAATCCTCTGTGAT 26519  
QY 1514 eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa 1534  
Db 26520 CCTGTCGACAGCAAAATTCGTGCTTAATCGCAGCATGAAGCAGCAGCCAGCAAAATTTGT 26579  
QY 1534 gIleLeuArgThrIleAlaValHisIleuGlnGlnLysPheArgGlyLysArgLeuMetI 1554  
Db 26580 CCAGCTGCGTACTATTTCGATTCATTTGCAACAAAAGTTCCGTGGCAAGCGTCTAATGAT 26639  
QY 1554 eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl 1574  
Db 26640 TGAGCAGCGTAATTGTTTCCAACTGCTCCGCTGTTCCATGCCGGGCTTCCAGCAGCGTGC 26699  
QY 1574 aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe 1594  
Db 26700 CCGCGCTTTATGGCTCGCAACAGATTCCAGGCCCTTGATGACACCCGAGATGATGACCT 26759  
QY 1594 uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrIleuIleAr 1614  
Db 26760 CATCCGCCAAGACGCGCGCCCAAGGTTATACAGCGTTTACTGGCGGGGCTTCTAATCCG 26819  
QY 1614 gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGl 1634  
Db 26820 ACAGCTCAGAGACCAACAGGTTCTTTGGATATCCGGAAGCGTATTGCTCAGCTCGGCA 26879  
QY 1634 nGluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuAr 1654  
Db 26880 AGAGGCAAAAGCGGTAAACTCTCTGCGCTGCAAGTCCAGAGGCGGTTCGTTCTTCG 26939  
QY 1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu----- 1669  
Db 26940 CCGAGCGCTTATCGCATCAGATCGGTTAGCTGCTTAAGTCGATTGGTAAAGTATCCAG 26999  
QY 1670 -----AspArgLeuSerAr 1674  
Db 27000 CTACGTTAGTCTTATTATACTAATCTGCTTAATAATFCCCGCAGAGTCTGCTTTTCGCG 27059  
QY 1674 gThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyI 1694  
Db 27060 CACTGTGCCACACCTGCTCATGTGTGTTTCGAGTTTCATGTCACGTTTTCGTATGCGAT 27119  
QY 1694 eMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgI 1714  
Db 27120 CATGCTCAGGCAATTCGATCAGAGTGGATAAGCAGCTTATAGAGCGCTGCAGCGGAT 27179  
QY 1714 eIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGlnGlyGly 1734  
Db 27180 CATCTTAATTTGGCCCGCTACATAGCACACCGTGAACACGTTCCAGAGCGCGGTTT 27239  
QY 1734 uValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluIlePheAsnTh 1754  
Db 27240 GGTCAACCATTCGCCAGATGTTATTTCGCTGGTGGTCCCAAGACAGTCAGATATTCAACAC 27299  
QY 1754 rIleuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770  
Db 27300 TTTGTGCAACCTCATTTGGGTATTTCGCCACTGTCTTAAAGCGAAAGGTAGCTAAATA 27359  
QY 1771 -----IleIleH 1773  
Db 27360 GCGAGAGATTAGGTTTATAGGAGTTTAAATCATATTCTTAATGCCCTTCCAGATCATTC 27419  
QY 1773 isAspTyrMetThrAsnProGluAlaIleTyrMetValArgGluThrLysLysLeuVal 1793  
Db 27420 ACGACTATGACCAACCCAGAGGCCATTTACATGTCGCGGAAACTAAGAAAGCTCGTGG 27479

QY 1793 labrglvsGluLysMetLysGlnAsnAlaAargLysProProMetThrSerGlyArgT 1813  
 Db 27480 CCGCAAGGAAAGATGAAGCAAAATGACGCAAGCCGCCCAATGACAAAGTGACGGT 27539  
 QY 1813 YrLysSerGlnLysIleAsnPhetheThrProCysSerLeuProSerLeuGluProAspPheG 1833  
 Db 27540 ATAAGACCCAAAGATAAACTTCACGCGGTTCCTCCCTGCCAGCGCTGAGCGGACATCG 27599  
 QY 1833 lyLeileArGtyrSerProTyrThrPheleSerSerValTyrAlaPheAspThrleL 1853  
 Db 27600 GAATCATCCGCTACAGTCCCTACAGTTATCTCGTCGCTTACGCTTCGATACGATTT 27659

QY 1853 euCysLysLeuGlnIleAspMetPhe 1861

Db 27660 TGTCAAGCTGCAGATGCATGTTT 27685

## RESULT 3

AC008205

LOCUS

DEFINITION AC008205 215711 bp DNA linear INV 24-FEB-2001

ACCESSION BACR33F18, complete sequence.

VERSION AC008205.7 GI:13122716

SOURCE HTG.

ORGANISM Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 215711)

Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gokeyne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,

Carlson,J.W., Center,A., Champ,M., Davenport,I.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,

Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ileguam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,

Phouanavong,S., Pittman,G.S., Furi,V., Richards,S., Scheeler,F.,

Stapleton,M., Strong,R., Swirskas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 96A-96B

Unpublished

2 (bases 1 to 215711)

Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,B., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

Direct Submission

Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 84-121, Berkeley, CA 94720, USA

On or before Feb 24, 2001 this sequence version replaced

gi:5701600, gi:7191021.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

Location/Qualifiers

1..215711

/organism="Drosophila melanogaster"

## FEATURES

source

/mol type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db xref="taxon.7227"  
 /chromosome="3R"  
 /map="96A-96B"  
 /clone="BACR33F18 (D764)"  
 /clone lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pBACE3.6)"

## ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 215711  
 Score: 9386.50 Matches: 1855  
 Percent Similarity: 96.32% Conservative: 4  
 Best Local Similarity: 96.11% Mismatches: 2  
 Query Match: 98.66% Indels: 69  
 DB: 3 Gaps: 3

US-09-914-698-1 (1-1861) x AC008205 (1-215711)

QY 1 MetGluLeuValTyrSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuLe 20  
 Db 40693 ATGGAGCTAGTGTGGAGCCCGCTCTTGAGGTGGCTTGCAGAGAGAGCGTGCAGCTAATA 40752

QY 21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40  
 Db 40753 GACAAACGCACTTCCGAAGAGAGTGTATGATCATCTCAAGTCCAAGAGCAACACGCG 40812

QY 41 Val----- 41  
 Db 40813 GTCAAGGTCAGTCAAAATAATGCAATTGCCACCCCTTTTACTTACATACCCCTCTTTTGCAT 40872

QY 42 -----LysAsnProArgLysPheProThrValGlyLysThrLe 54  
 Db 40873 GTATCTCTTACCTTTTACCTGGAAGAACCCGCGCAAAATTTCTTACTGTCCGCAAGACCCCT 40932

QY 54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValValSerAlaAlaVa 74  
 Db 40933 GCAGCTGAATCCGCGAGAGAGTGGCAAGACAATGAAAGCGTGGTATCGCTGTGT 40992

QY 74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94  
 Db 40993 GCAGCAAAAGAGCGCATGTCTGCAGCAGCAGCGCTCCCTCCAAGCAGACATGCGCAGT 41052

QY 94 lThrAlaProSerArgProAlaAlaTtpAlaHisProProGlnAlaProLeuValG1 114  
 Db 41053 GACTGCTCTTTCGCGTCCCGCTGCATGGGCAATCCACCTCCAGAGGCTCCTCTTGTGCA 41112

QY 114 uLysAsnValTyrLysThrProGlnGluProValTyrIleSerProGlnProArgSe 134  
 Db 41113 GAAGNATGTATACAAGACTCCACAAGAAGAGCCCGGTATACATATCACCACAGCCTCGCAG 41172

QY 134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154  
 Db 41173 TCTTAAGGAAATCTTAAGCCCTATGACGCCAGGAAACCTACTCGACGCTGATGCAATCT 41232

QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174  
 Db 41233 CGGATTACACCTCTCACCGAAACCCGTGGCAAGGACAGCTACCATTTTCCCGGACAA 41292

QY 174 nLeuAlaAlaTyrProThrProThrLeuLysGlyLysValLysSerCysAlaAsnAspMe 194  
 Db 41293 TCTGGCAGCTGGCCCAACCAACACACTTAAAGGGAATGAAATCATGTGCTAATGATAT 41352

QY 194 tArgProArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214  
 Db 41353 GCGGCGCGTCCGAATCACTCCAGATGATCTAGAGATCAGCGTGCACAAACAAACGTT 41412

QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArg11 234  
 Db 41413 CGATGAAGCAATCCGAGACTATCAATATTTGGTGGACACCTTGGAGCTGTCCAGGAT 41472

QY 234 eAspGlycInProHisThrProLeuAsnLysThrThrThrIleValHisAlaThrHisTh 254  
Db 41473 CGATGGACAACCGCATACGCCCTTAATAAGACAACAACCATTTGTGCATGCCACGACAC 41532  
QY 254 rArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThrThrLys 274  
Db 41533 CAGAGCTCTGGCCCTGTATTAATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 41592  
QY 274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294  
Db 41593 GAGCGCCATACACGACCTGAAGAGGACATTAAGTTGGTGGTTTCACTTACGAAAGTA 41652  
QY 294 rSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrrAlaIleGlnI 314  
Db 41653 TTCCGAGTCCATGAAGATTTGTCTTATCGGCCCAAACTAAGATGATGCCATCAAGG 41712  
QY 314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrrGI 334  
Db 41713 GTCTATGCCCTAATCTAATGAAATGAATCCCGCTCGATCGAACAGATATATACCA 41772  
QY 334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaSe 354  
Db 41773 GGAGCAGCAGATCCAGATTAAGGCCAAAGACTTGAATAGCTCTCTAGTAGCGAGGTAG 41832  
QY 354 rLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerAr 374  
Db 41833 TTTGGCCGCGCAGCAGGAGTTTCTAATCAACACAGTAGGATCTCTAGTAGCGAGGTAG 41892  
QY 374 gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi 394  
Db 41893 TTTTAACTCCATGAAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 41952  
QY 394 sLysArgSerHisGlnLeuSerPheSerAspAlaProSerAsnGlnSerLeuTyrrAr 414  
Db 41953 CAAGCGCGCTCTCATGAGTTGAGTTTTCGGATGACCTAGCAACGAATCATTTGTACCG 42012  
QY 414 qAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr 434  
Db 42013 CAATGAACCTGTAGCAATTTCCCTCTCTAAAGCAACGGGTTGAGGACACTTACTGTC 42072  
QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHi 454  
Db 42073 CAGAAGTGCAGCGCGGCAATGCATCTGCAAGAGCAGTAGTGCCTCAGCGCTGGCCACA 42132  
QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
Db 42133 CGCCCAATCCCAAGAGTTTAACTAGCTAGCACAAACCATGTCTACTGATGAAGAGCCG 42192  
QY 474 rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyrrAspSerGluLe 494  
Db 42193 ACCACGAAAGTCAAGGACACTAGCAATTCAGCTTCCGTCAGCTTATGACTCGAGCT 42252  
QY 494 uTyrrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThrIleAspProPh 514  
Db 42253 GTATATGACAGCTGCATATCAACCGGATTCATTTGCGAGCACTACGCAATTCATTCATT 42312  
QY 514 eLeuAlaSerThrMetTyrrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy 534  
Db 42313 TCTGGCATCTACCATGTATTTGGATGAACAGGCTGTGATCGTCATCAAGCTGACTTTAA 42372  
QY 534 sLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnLys 554  
Db 42373 AAAAGTGTAAATGCCCTTGTCTCATACCGCTGAGCTGAGCGAGATTAATAACAA 42432  
QY 554 sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValValAlaProTh 574  
Db 42433 AATAGACGTTGTAGCTGTATTAACAGAGTGGCCAAACAGAGCTCTGTGGTGGCTCCAC 42492  
QY 574 rLysGluGlnGlnSerMetAsnTyrrLeuThrLysTyrrArgLeuGluThrLeuArgLysAl 594  
Db 42493 CAAGGAGGAGCAGTCTATGAACCTACCTAAACGAAATACCGCTGAGAGCGCTTCGTAAGG 42552  
QY 594 aAlaValGluLeuPheSerGluGlnMetArgLeuProCysSerLysValAlaValTy 614

Db 42553 GCCTGTGGAGCTCTTCTTCACTGAGCAGATGCGCTGCATCTCCAAAGTGGCGGTATA 42612  
QY 614 rValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMe 634  
Db 42613 TGTCAACAGACAGCTCTGGCATCCGTAGCATCGTAATCTTCACTAGACGTAGTTAT 42672  
QY 634 tGlnArgThrIleLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLe 654  
Db 42673 GCAAGCGCACCATCTTGTAGCTGTCTTGTCTCAATCCCTTGGCTGGCGCTGGACT 42732  
QY 654 uGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe 674  
Db 42733 GGAAGTGGTCTTTGGCGAGAGATCCAGATGAGTCTAATCGAGACATTTAGGCTCAG 42792  
QY 674 rThrPheIleLeuAsnArgLeuPheAsnLysCysGluGluGlnArgTyrrSerLysAl 694  
Db 42793 CACCTTTATCTCAATCGCTTGTTCGGAATAAGTGTGAGGAGCAGAGTACAGCAAGG 42852  
QY 694 aTyrrThrLeuThrGluGluTyrrAlaGluThrIleLysLysHisSerLeuGlnLysIleLe 714  
Db 42853 ATACACACTCACGAGAGTACGCGAGACCATTAAGAAGCACTCATTTGCAGAAATCT 42912  
QY 714 uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCy 734  
Db 42913 CTTTGTGTCTTCTTCTCGATCAAGCTAAGCAGAGGAGCATCTGTCAGACACATCCCTG 42972  
QY 734 sLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSe 754  
Db 42973 TTTTGTGTAAAGTAGTCCACATAAAGAGACCAAGGATATTTCTGTGCGCTTCCTGTC 43032  
QY 754 rGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrrValLe 774  
Db 43033 GGAGCTGTCTCGCAACATTTGGTATTTACGCGGAACTTCTGCGCTGGGCTAGCTTCT 43092  
QY 774 uGlnHisArgGlnThrPheLeuAspGluPheAspTyrrAlaPheAsnAsnLeuAlaValAs 794  
Db 43093 ACAGCACCGCCAAACATTTTGGACGAGTTCGATATGCTTCAACAACTTGGCTGTGGA 43152  
QY 794 pLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAspLe 814  
Db 43153 CCTAAGAGATGCGCTGAGACTAACCCGAGTATGAGGTAATTTTACTACGCGATGATCT 43212  
QY 814 uThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLe 834  
Db 43213 AACCCGCGAGTTAAGGTTGCCAGCCATCTCTGCGCTTACGCGGATCTTCAATGTAAAGCT 43272  
QY 834 uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAs 854  
Db 43273 AGCTCTGGCGCACCTTGTGTGAAGCCAACTTCAGCTAGCGCGGACATCGCGGCCAAGA 43332  
QY 854 pIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrrLysPh 874  
Db 43333 CATCTGTGACGACATCTGTGAGAACCGCTTTCCTGCTCTGGCAACTTATTTACAAAT 43392  
QY 874 eArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTr 894  
Db 43393 CGCTCGCCCAAGTTTCAATGCGGCGCCACGCTGTCTCCAGAAATGGTGGCGCGTCACTG 43452  
QY 894 pLeuHisValIleGlnArgArgIleArgHisLysGluLeuMetArgArgHisArgAl 914  
Db 43453 GCTGACGCTGTGTATTCAGCGCTCGCATTCGCCACAAAGAGCTTATGCGTCCGCCCGGCG 43512  
QY 914 aAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrrValLysLeuPh 934  
Db 43513 CGCTACTGTCTTCACTGAGCGGTTCCTGGGCCACCATGAGATGAGATGAGTGAAGTGT 43572  
QY 934 eLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrrLeuAl 954  
Db 43573 TAAGACGGAACGCACTCAGGCGCGCAATAATTCGCAAAAGTTTCAACCGCTCGGTATTTGGC 43632  
QY 954 aGlnLysGlnLeuTyrrGlnSerTyrrHisSerIleIleThrIleGlnArgTyrrTrpArgAl 974

43633	DB	CCGAAAGCAGCTATATACAGAGCTATACACAGTATTATACCAATCCAGCCCTGGTGGCGAGC	43694
974	QY	aGlnGlnLeuGluYargGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh	994
43693	DB	CCAAACAATGGGAAGCAGCACCGCAGCGGTTTGTGGAGCTCCAGAGGGCTGCGATCTT	43752
994	QY	eLeuGlnArgIleTrrpArgArgIeuPheAlaIysIysLeuLeuAlaAlaGluThr	1014
43753	DB	TCATTACGGGATCTGGCGGCGACGACTCTTTGTCGAAAAAATATTGGGGGGCGGGAAC	43812
1014	QY	rAlaArgLeuGlnArgSerGlnIysGlnAlaAlaIaIaSerTyrIleGlnMetGlnTr	1034
43813	DB	AGCCAGACTTCACGCGATCGCAAAAACAACGACAGCTGCTAGTTATATTCAATGCAATG	43872
1034	QY	pArgThrTyrGlnLeuGluYargIleGlnArgHisGluPheLeuArgGlnArgAspLeuI	1054
43873	DB	GCGAAGTATCAGCTGGGCGAGAAATTCAGCGACAGCAGTCTCTGCGCCAGAGGACCTCAT	43932
1054	QY	eMetPheValGlnArgMetArgSerIysTrpSerMetLeuGluGlnArgLysGluPh	1074
43933	DB	CATGTTTGTTTCAGCGCAGGATGCGAAGCGAGTGAGTATGCTGGAGCAGCGCAGAGGTT	43992
1074	QY	eGlnGlnLeuLysArgAlaAlaIleasnIleGlnArgGlnArgTrrpArgAlaLysLeuSerMe	1094
43993	DB	CCAGCAACTAAAGCGTGCAGCTATAAATATCAACAACGCTGGCGAGCGAAGCTTTCAAT	44052
1094	QY	tArgLysCysAsnAlaAspTrrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTy	1114
44053	DB	GAGAAAGTCGAACGCCTGATTATTGGCATCTCGTTCCACGCTTCTTAAAGTTCCAGGCTTA	44112
1114	QY	rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTrrTrrSerLeuArgLysAs	1134
44113	DB	CAGGAAAGCCACAATCCAGATGAGATAGATCGTAATCACTACTATTCCCTGGGAAAAA	44172
1134	QY	nValIleCysLeuGlnGlnArgLeuArgAlaIleMetIysMetArgGluGlnArgGluAs	1154
44173	DB	TGTTATTCTGCTGCACACGCACTGAGGGCCATCATGAAATCGCGCAACAGAGGGAAAA	44232
1154	QY	nTrrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTrrArgMetArgGlnGl	1174
44233	DB	TTTATCTGAGGCTCGGAAATGCTTCAATACATAGTTCAAAAACGCTACCGCATCGCTCAACA	44292
1174	QY	nMetIleGlnAspArgAsnAlaTrrLeuArgThrArgLysCysIleIleAsnValGlnAr	1194
44293	DB	AATGATCCAGGATAGAAATGCATATTTAAGAACCCGCAAAATGTATCATCAATGTCACAG	44352
1194	QY	qArgTrrpArgAlaThrLeuGlnMetArgGluArgLysAsnTrrLeuHisLeuGlnTh	1214
44353	DB	GCCTTGAGAGCCACTCTCGCATGGCTCGAAGAAGAACTACCTTTCATCTCCAAAC	44412
1214	QY	rThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl	1234
44413	DB	GACAACCAACGAATTCAAATCAAGTTCGTCGCCAAGCGGAAATGAAAAAGCAAGAGC	44472
1234	QY	aGluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLe	1254
44473	DB	CGAGTTCTCTCAGCTGAATAAGGTTACCCCTGTAGTCCAGAAACGTCGCGCGGCTTGCT	44532
1254	QY	uGlnMetArgLysGluArgGlnGluTrrLeuHisLeuArgGluValThrIleLysLeuGl	1274
44533	DB	GCAGATCGGAAAGGAGCGCCAGGATACCTACATCTGGCGGAGGTTCATCAATAAGCTGCA	44592
1274	QY	nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTrrArgGlyThrGl	1294
44593	DB	GCGCAGATTCCATGCTCAAAAATCAATGCGGTTCTATGGAGCCAAAGTACC CGCGCACCCA	44652
1294	QY	nAlaAlaValSerCysLeuGlnMetHisTrrpArgAsnHisLeuLeuArgLysArgGluAr	1314
44653	DB	GGCTGCTGTGAGGTGCTGCAGATGCAATTGGCGTAAATCATCTCTCTTAGAAAACGGAGAG	44712
1314	QY	gAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTrrArgAlaAr	1334
44713	DB	GAACAGTTTCTTCAACTCGGTCACAGACGATCAACACTGACGACACGGTACCGAGTCTG	44772

Qy	1334	gLeuAsnMetIleIysGlnLeuIysSerTyrAlaGlnLeuIysGlnAlaAilaIleThrIi	1354
Db	44773	TCTGAATATGATCAAGCAGTTGAAGAGTTACGCCAGCTGAACACAGCGAGCTATTACCAT	44832
Qy	1354	eGlnThrArgTyrArgAlaIalysAlaMetGlnIysGlnValIleValIeuTyrGlnIysG1	1374
Db	44833	TCAAACCCGATATAGAGCCAGAGGCAATGCMAAAGCAGGTGGTCTTGTACCAAAAGCA	44892
Qy	1374	nArgGluAlaIleIleIysValGlnArgArgTyrArgGlyAsnLeuGluMetArgIysG1	1394
Db	44893	AAGGGAGGCCATTATCAAAGTGCACGACGATACCCGGCAATCTGGAGATGAGGAAGCA	44952
Qy	1394	nIleGluValTyrGlnIysGlnArgGlnAlaValIleArgLeuGlnIysTyrTrpArgSe	1414
Db	44953	GATTGAAGTCTACCAAAACAGCCCGCAGCTCATCCGTTGCAGAAATGCTGGCGCAG	45012
Qy	1414	rIleArgAspMetArgLeuCysIysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe	1434
Db	45013	TATACGCGCATCGCGCTGTGCAAGCGGGCTACCGAAGATTCGACTCAGTTCATTGAG	45072
Qy	1434	rIleGlnArgGlyTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe	1454
Db	45073	CATTCAACGCAAGTGGCGGGCCACAGTGCAGGCTCGTCGCCAACGGGAGATCTTCTTAAG	45132
Qy	1454	rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgG1	1474
Db	45133	CACCATCCGCAAGTCGACTTATGACAGCATTTATCAGAGCAACTTACTGTGATCGTCA	45192
Qy	1474	nGlnArgArgGluPheGluMetIysArgArgAlaAlaValIleGlnArgArgPheAr	1494
Db	45193	GCAGCGCAGGGAATTCAGATGAAGCGAAGGGCTCGGTAGTGATCCAGCGCCGCTTCG	45252
Qy	1494	gAlaArgCysAlaMetLeuIysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValI1	1514
Db	45253	TGCTCGCTGTGCTATGCTTAAAGCGCAGACAGGATTATCCAATTAATCCAATCTCTGTGAT	45312
Qy	1514	eLeuValGlnArgLysPheArgAlaAsnArgSerMetIysGlnAlaArgGlnGluPheVa	1534
Db	45313	CCTGCTGCAGCGCAATTCGCTGCTAAATCGCAGCATGAAGCAGGCAAGCCAGGAATTTGT	45372
Qy	1534	lGlnLeuArgThrIleAlaValIHisLeuGlnGlnIysPheArgGlyIysArgLeuMetI1	1554
Db	45373	CCAGCTCGCTGCTATTTGCTATTCAGTCCATTTGCACAAAGATTCGCTGCGCAAGCGTCTAATGAT	45432
Qy	1554	eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl	1574
Db	45433	TGAGCAGCGTAATTTGTTTCCAACTGCTCGCTGTTCCATGCGCGGCTCCAGGCACGTCG	45492
Qy	1574	aArgGlyPheMetAlaArgIysArgPheGlnAlaIleuMetThrProGluMetMetAspLe	1594
Db	45493	CCGCGGCTTTATGGCTCGCAAAACGATTCAGGCGCTGATGACACCCGAGATGATGGACCT	45552
Qy	1594	uIleArgGlnLysArgAlaAlaIalysValIleGlnArgTyrTrpArgGlyTyrLeuIleAr	1614
Db	45553	CATCCGCCAAGACGCGCCGCCAAGGTATACAGCGTTACTGGCGGGCTATCTAATCCG	45612
Qy	1614	gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgIysArgIleAlaGlnLeuArgG1	1634
Db	45613	ACAGCTCGAAGACACCAAGGCTCTCTGTGATATCCGGAAGCGATATGTCTAGCTGCGGCA	45672
Qy	1634	nGluAlaLysAlaValAsnSerValArgCysLysValGlnGlnAlaValaArgPheLeuAr	1654
Db	45673	AGAGGCCAAAGGCGGTAACCTGTGTGCGTGCMAAGTCCAGGAGGCGGTGCGTTTCTCTTCG	45732
Qy	1654	gGlyArgPheIleAlaSerAspAlaIleuAlaValLeuSerGlnLeu-----1669	45792
Db	45733	CGGACGCTTTATCGCATCAGATCGGTATGACGTGCTAAGTGTGATTTGGGTAGTGTATCCAG	45792
Qy	1670	-----AspArgLeuSerAr1674	45852
Db	45793	CTACGTTAGTCCCTATTATAACTAATCTCGCTTAATATATCCCGGACAGATCGTCTTTCGCG	45852







QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174  
 Db 44803 GCGATTCCAGACCTCTCCAGGAACCCGGTGGCAAGAGCAAGCTACCAATTTCCCGGACAA 44862  
 QY 174 nLeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMe 194  
 Db 44863 TCTGGCAGCCTGGCCACACCAACACTTAAAGGGAATGTAAATCATGTGCTAATGATAT 44922  
 QY 194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214  
 Db 44923 GCGCCCGGTGGGAATCACTCCAGATGATCAGAGATCAGAGATCAGCCGCAACAAACGAT 44982  
 QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArgI 234  
 Db 44983 CGATGTAAGACATTCGAGACTATCATATTTGGTGGACACCTTGAGCTGCTCCAGGAT 45042  
 QY 234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisIsth 254  
 Db 45043 CGATGGACAACCGCATACGCCCTTAAATAAGACAACAACCAATTTGTGCATGCCAGCACAC 45102  
 QY 254 rArgAlaLeuAlaCysIleHisGluGluGlyProSerProArgThrProThrLy 274  
 Db 45103 CAGAGCTCTGCGCTGTATTCATGAGGAGGAGGACCAAGTCCACTAGGACGCCCGACGAA 45162  
 QY 274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294  
 Db 45163 GAGCGCCATACAGACTGAAGAGGACANTAGTTGGTGGTTTACCCCTTACGAAGTA 45222  
 QY 294 rSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyAlaIleGlnGl 314  
 Db 45223 TTCGAGTCCATGAAGAATTTGTCACTTCATCGCCACAAACAACTAAGTATGCCATTCAGG 45282  
 QY 314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyTrGI 334  
 Db 45283 GTCTATGCTCTAAATGAATGAATAATCCGCTCGATCGAACAAGATAGATATTACCA 45342  
 QY 334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaIse 354  
 Db 45343 GGAGCAGCAGATCCAGATAAGGCCAAAGACTTGATAGCTCTCTAGTAGCGAGGCTAG 45402  
 QY 354 rLeuAlaGlyGlnGlnPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374  
 Db 45403 TTTGGCGGCGCAGCAGAGTTTCTAATCAACACAGTGAATCCTCGCTCAGTCCAGTGC 45462  
 QY 374 gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi 394  
 Db 45463 TTTTAACTCCATGAAGTAGGTGGGAAGTGGGTGAAGGGGAAGTCCGGTGAAGGAATCCTCA 45522  
 QY 394 sLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyTrAr 414  
 Db 45523 CAAGCGCGCTCTCATGAGTTGAGTTTTCGGATGCACTAGCAACGAATCATTTGTACCG 45582  
 QY 414 gAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr 434  
 Db 45583 CAATGAACATGTAGCCATTTCCCTCTCTAAAAAGCAACGGGTGAGGACACTACTCTGCC 45642  
 QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHi 454  
 Db 45643 CAGAAGTGAGCGCGCGGCAAAATGCATCTGCAGAGAAGCAGTAGTGGCCACGCCCTGGCCACA 45702  
 QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
 Db 45703 CGCCCAATCCAAGAAGTTAAGCTAGCACAAACCATTTCTCATGATGAAGAGCCGCCAC 45762  
 QY 474 rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyAspSerGluLe 494  
 Db 45763 ACCACGAAAAGTCAGGACACTAGCATTCAGCCCTCCGTCAAGCTCTATGACTCGGAGCT 45822  
 QY 494 uTyTrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrIleAspProPh 514  
 Db 45823 GTATATGCAGAGTGATCAACCCCGGATCCATTTGAGCAACTTACGACAAATGATCCATT 45882

QY 514 eLeuAlaSerThrMetTyTrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy 534  
 Db 45883 TCTGGCATCTACCATGATTATTGGATGAACAGGCTGTGGATCGTCATCAAGCTGACTTTAA 45942  
 QY 534 sLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsnLy 554  
 Db 45943 AAAGTGGTTAAATGCCCCCTGCTCTCCATACCGCTGACCTGGACGACAGTTTAAATACAA 46002  
 QY 554 sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValAlaLPrOth 574  
 Db 46003 AATAGACGTTGGTAAGCTGTTTAAACGAGGTGCGCAACAAAGAGCTCGTGGTGGCTCCAC 46062  
 QY 574 rLysGluGluGlnSerMetAsnTyTrLeuThrLysTyArgLeuGluThrLeuArgLysAl 594  
 Db 46063 CAAGGAGGACAGCTCTATGAACCTACCTAACGAATACCGCTCGAGACGCTTGTAAAGGC 46122  
 QY 594 aAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTy 614  
 Db 46123 GGCTGTGGAGCTCTTCTTTCAGTGAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATA 46182  
 QY 614 rValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMe 634  
 Db 46183 TGTCAACAAGCAAGCTCTGCGCATCCGTAGCGATCGTAATCTTCACTAGACGTAGTTAT 46242  
 QY 634 tGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLe 654  
 Db 46243 GCMAACGACCATACTTGGAGCTGCTGCTTCTTCAATCCCTTTGGTGGCTCGGCTTGGACT 46302  
 QY 654 uGluValValPheGlyGlyLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe 674  
 Db 46303 GGAAGTGGTCTTTGGCGAGAGATCCAGATGCACTCTAATCGAGACATTTGAGCCCTCAG 46362  
 QY 674 rThrPheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyTrSerLysAl 694  
 Db 46363 CACCTTTATCTCAATCGCTTGTTCGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC 46422  
 QY 694 aTyTrThrLeuThrGluGluTyAlaGluThrIleLysLysHisSerLeuGlnLysIleLe 714  
 Db 46423 ATACACACTCACCAAGAGTACCGGAGACCATTAAGAAGCACTCATTCAGAAAAATCCT 46482  
 QY 714 uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCy 734  
 Db 46483 CTTTTTGTCTTCTTCGATCAAGCTAAGCAGAGCGCATCGTCAAGCACATTCCTTG 46542  
 QY 734 sLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSe 754  
 Db 46543 TTTGTTTGTAAAGATCGCCACATAAAGAGACCAAGGATATTCTGCTGCGCTTCTCGTC 46602  
 QY 754 rGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyTrValle 774  
 Db 46603 GGAGCTGCTCGCAACATTTGGTGTATATTACGGGGAACCTTCGTCGCTCGGCTTACGTTCT 46662  
 QY 774 uGlnHisArgGlnThrPheLeuAspGluPheAspTyTrAlaPheAsnAsnLeuAlaValAs 794  
 Db 46663 ACAGCACCGCCCAACATTTTGGACGAGTTGCTATGCTTCAACAACTTGGCTGGA 46722  
 QY 794 pLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAsple 814  
 Db 46723 CCTAAGAGATGGCGTGAGACTAACCGAGTCTATGAGGTAATTTTACTACGCGATGATCT 46782  
 QY 814 uThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLe 834  
 Db 46783 AACCGCCCAAGTTAAGGTGCCAGCATCTCTCGCTTTCAGCGGATCTTCAATGTAAAGCT 46842  
 QY 834 uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAs 854  
 Db 46843 AGCTCTGGGCGCACCTTGTGTAGAGCCAACTTCCAGCTAGCGCGGACATCGCCGCCAAGA 46902  
 QY 854 pIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIleTyTrLysPh 874  
 Db 46903 CATCGTTGACGAGACATCGTGAGAGACGCTTTCCCTGCTCTGCAACTTATTATTACAAAT 46962  
 QY 874 eArgSerProLysPheHisAlaAlaThrValLeuGlnLysTyTrpTrpArgArgHisTr 894



Db	46963	CCGCTCGCCCAAGTTTCATGCGCGGCCACCGTGCTCCAGAAATGGTGGCCGCTCACTG	47022	Db	48043	CGAGTTTCTTCAGCTGAAAAAGGTACCTCTGTAGTCCAGAAACGTCGGCGGGCTTGCT	48102
Qy	894	pLeuHisValVallleGlnArgArgilleArgHisLysGluLeuMetArgArgHisArgAl	914	Qy	1254	uGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrilleLysLeuGl	1274
Db	47023	GCTGCAGCTGTATTCAGCGCTGCATTCGCCACAAAGAGCTTATGCGTCGCCACCGGGC	47082	Db	48103	GCAGATCGAAAGAGCGCCCGAGGATACCTACATCTGCGCGAGGTACCATTAAGCTGCA	48162
Qy	914	aAlaThrVallleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPh	934	Qy	1274	nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl	1294
Db	47083	CGCTACTGTCTATTGAGCCCGTTCGCTGGGCCACAGATGAGAAAGTACGTGAAGTGTGT	47142	Db	48163	CGCGAGATTCCATGCTCAAAAATCAATCGCGTTCATGCGAGCAAGTACCGCGCACCA	48222
Qy	934	eLysThrGluArgThrGlnAlaAlaIleleLeuGlnLysPheThrArgArgTyrLeuAl	954	Qy	1294	nAlaAlaValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluAr	1314
Db	47143	TAAGACGAAACGCACTCAGCGCGCAATCTTCGAAAAGTTCACCCGTCGGTATTGGC	47202	Db	48223	GGCTGCTGTAGCTGCTGCGAGATGCATTGGCGTAATCATCTCTTAGAAACGGGAGAG	48282
Qy	954	aGlnLysGlnLeuTyrGlnSerTyrHisSerIleleThrleGlnArgTrpTrpArgAl	974	Qy	1314	gAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrleLeuGlnArgArgTyrArgAlaAr	1334
Db	47203	CCAGAAACGAGCTATATCAGAGCTATCAGATTATATCACCATTCCAGCGCTGCTGGCGAGC	47262	Db	48343	TCTGAATATGATCAAGCAGTTTGAAGAGTTACGCCAGCTGAAACAGGCGAGCTATTACCAT	48402
Qy	974	aGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh	994	Qy	1354	eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValVallleuTyrGlnLysGl	1374
Db	47263	CCAAACAACCTGGGAAGGACGACCGCCAGCGGTTTGTGGAGCTCCGAGAGGCTGCGATCTT	47322	Db	48403	TCAAAACCGATATAGAGCAAGAGGCAATGCAAAAGCAGGTGGTCTTTGTACCAAAAGCA	48462
Qy	994	eLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluTh	1014	Qy	1374	nArgGluAlaIlelleLysValGlnArgTyrArgGlyVAsnLeuGluMetArgLysGl	1394
Db	47323	TCTTCAGCGGATCTGCGCGGACGACTCTTTGGCCAAAACCTATTGGCGCGCGGAAAC	47382	Db	48463	AAGGAGGCCAATTATCAAGHGCAACGACGATACGCGGCAATCTGGAGATGAGGAAGCA	48522
Qy	1014	rAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTr	1034	Qy	1394	nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSe	1414
Db	47383	AGCCAGACTTCAGCGATCGCAAAAAACAACAGCAGCTGCTAGTTATATTCAAATGCAATG	47442	Db	48523	GATTGAAGTCTACCAAAAACAGCGCCAGCGCATCTCCGCTTGCAAAAATGGTGGCGAG	48582
Qy	1034	pArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIl	1054	Qy	1414	rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe	1434
Db	47443	CGGAAGCTATCAGCTGGGCGAATTCAGCGACAGCGTTCCTGGCCGAGAGGACCTCAT	47502	Db	48583	TATACGCGACATGCGGCTGTCAAAGCGGGCTACCGAAGGATTCGACTCAGTTTCATTGAG	48642
Qy	1054	eMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPh	1074	Qy	1434	rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe	1454
Db	47503	CATGTTTGTTCAGCGCAGGATCGGAAGCAAGTGGAGTATGCTGGAGCAGCGCAAGAGTT	47562	Db	48643	CATTCAACGCAAGTGGCGGCGCACAGTGCAGGCTGTCGCCACGCGAGATCTTCTTAAG	48702
Qy	1074	eGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMe	1094	Qy	1454	rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl	1474
Db	47563	CCAGCAACTAAAGCGTCGAGCTATATAATATCCAAACGCTGGCGAGCGAAGCTTTCAAT	47622	Db	48703	CACCATCGCAAAAGTGCAGCTTATCGAGCATTTATCAGAGCAACTTACTGATCGGTCA	48762
Qy	1094	tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAlaTy	1114	Qy	1474	nGlnArgArgGluPheGluMetLysArgArgAlaAlaValVallleGlnArgArgPhear	1494
Db	47623	GAGAAAGTGCACCGCTGATATTTCGCACTTCGTTCCAGCGTTCITAAAGTTACGGCTTA	47682	Db	48763	GCAGCGCAGGGAATTCGAGATGAAGCGAAGGCTGCCGTAGTATCCAGCGCGGTTTCG	48822
Qy	1114	rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs	1134	Qy	1494	gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl	1514
Db	47683	CAGGAAGCCCAATCCAGATGAGATAGATCGTAACTACTACTATTTCCTCGCAAAAAA	47742	Db	48823	TGCTCGCTGTGCTATGCTTAAGCGCAGACAGGATTAACCAATTAATCCATCTCTGTGAT	48882
Qy	1134	nVallleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAs	1154	Qy	1514	eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa	1534
Db	47743	TGTTATCTGCTGCAACAGCACTGAGGCGCATCATGAAATGCGCAACAGAGGGAATA	47802	Db	48883	CCTGTGTGCGCGCAAAATTCGCTGCTAAATCGCAGCATGAAGCAGGCGCACGCGAAATTTGT	48942
Qy	1154	nTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGl	1174	Qy	1534	lGlnLeuArgThrIleAlaValHisLeuGlnLysPheArgGlyLysArgLeuMetIl	1554
Db	47803	TTATCTGAGCGTGCAGAAATGCTTCAATACTAGTTCAAAACGCTACCGCATCGCTCAACA	47862	Db	48943	CCAGCTGCGTACTATTTGAGTCCATTTGCCAAAAGTTCCGTGGCAGCGCTTAATGAT	49002
Qy	1174	nMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleleAsnValGlnAr	1194	Qy	1554	eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl	1574
Db	47863	AATGATCCAGGATAGAAATGCATATTTAAGAACCCGCAATGTATCATCAATGTCAGAG	47922	Db	49003	TGAGCAGCGTAATGCTTTTCCAACTGCTCGCTGTTTCCATCGCGGGCTTCCAGGCGCTGC	49062
Qy	1194	gArgTrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnTh	1214	Qy	1574	aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspie	1594
Db	47923	GGCTGGAGGACCTCTGCAGATCGTCGAGAAAGAGAACTACTCTTCATCTCCAAAC	47982	Db	49063	CGCGGCTTTATGGCTCGCAACGATTCAGGCCCTGATGACACCGCAATGATGACCTT	49122
Qy	1214	rThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl	1234	Qy	1594	uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleAr	1614
Db	47983	GACAACCAACCAATTCAAATCAAGTTCCGTGCCAAGCGCGAAATGAAAAAGCAAGAGC	48042	Db	49123	CATCGCCAGAAAGCGCGGCCCGCAGGTATATACAGGCTTACTGCGGGGCTTAICTAATCCG	49182
Qy	1234	aGluPheLeuGlnLeuLysValThrLeuValValGlnLysArgArgArgAlaLeuLe	1254				

```

1614 gAArgGlnlySHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgG1 1634
49183 ACAGAGTCAGAGCACCAGGTCCTCTGGATATCCGGAAGCGTATTGCTCAGTGGCGCA 49242
1634 nGUAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuAr 1654
49243 AGAGGCAAGAGCGGTAACTCTGGCGCTGCAAGTCAGAGGCGGTGGTTCTCTTCG 49302
1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu----- 1669
49303 CGAGCGCTTTATCGCATCGATCGGTAGCAGTGTAAAGTCGATGGTAAAGTCATCCAG 49362
1670 -----AspArgLeuSerAr 1674
49363 CTACGTTAGTCCTTATTATAACTAATCTGCTTAATAATCCCGGAGATCGCTTCGCG 49422
1674 gThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTrpGlyI1 1694
49423 CACTGTGCCACACCTGCTCATGTGTGTTCGGAGTTCATGTCCACGTTTTCGCTATGCCAT 49482
1694 eMetaLagInAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgI1 1714
49483 CATGGCTCAGGCCATTCGATCAGAGTGGATAGACAGCCTTATAGAGCGCTGCGACCGGAT 49542
1714 eIleLeuAsnLeuAlaArgTyzAsnSerThrThrValAsnThrPheGlnGluGlyGlyLe 1734
49543 CATCTTAATTTGGCCGCTACATAGCACCCAGCGTGACACGTTCCAGAGGCGCGTTT 49602
1734 vAlaThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluIlePheAsnTh 1754
49603 GGTCAACCATTTGCCAGATGTTATTGGCTGTGGTGGCAGACAGACAGTGCATATTCACAC 49662
1754 rLeuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770
49663 TTTGTGCACCTTCATTTGGTATTTCGCCACTGCTCTAAAGAGCGAAGAGTAGTAGTAAATA 49722
1771 -----IleIleH 1773
49723 GCGAGAGGATTAGTTTACGAGTTTAAATCATATTCCTTAATGCCTCCAGATCATTC 49782
1773 iAspTyTrMetThrAsnProGluAlaIleTyMetValArgGluThrLysLysLeuValA 1793
49783 AGCACTATATGACCAACCCAGAGGCGCATTTACATGTGTGGCGGAACTAAGAGAGCTGTGG 49842
1793 laArgLysGlnLysMetLysGlnAsnAlaAcLysProProMetThrSerGlyArgT 1813
49843 CCGCAAGGAAAGATGAGCAAAAATGACAGCAGCGCGCCCAATGACAAAGTGCAGCGT 49902
1813 yrLysSerGlnLysIleAsnPheThrProCysSerLeuProSerLeuGluProAspPheG 1833
49903 ATAGAGCCAAAAGATAAACTTCACGCGCTGTTCTCTGCCAGCTGGAGCGCGACTTCG 49962
1833 lyIleIleArgTySerProTyThrPheIleSerSerValTyAlaPheAspThrIleL 1853
49963 GAATCATCGCTACAGTCCCTACACGTTTATCGTCGTTTACGCCCTTCGATACGATT 50022
1853 euCysLysLeuGlnIleAspMetPhe 1861
50023 TGTGCAAGCTGCAGATCGACATGTTT 50048

```

```

RESULT 5
LOCUS AY094825
DEFINITION Drosophila melanogaster LD39479 full insert cDNA.
ACCESSION AY094825
VERSION AY094825.1 GI:20151636
KEYWORDS FLI CDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

```

## REFERENCE

## AUTHORS

1 (bases 1 to 3486)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
 Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,  
 George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G.,  
 Miranda, A., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,  
 Patel, S., Phouanavong, S., Wan, X., Yu, C., Lewis, S. E., Rubin, G. M.  
 and Celniker, S.

## Direct Submission

Submitted (03-APR-2002) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA

## Sequence submitted by:

Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory

## Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our Web site  
 (<http://fruitfly.berkeley.edu>) or send email to  
[cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

## Source

1..3486  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /map="96A23-96A23"  
 1..3486

## gene

/gene="asp"  
 /note="alignment with genomic scaffold AE003749"  
 /db\_xref="FLYBASE:FBgn0000140"  
 82..2715

## CDS

/gene="asp"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="LD39479p"  
 /protein\_id="AA011178.1"  
 /db\_xref="GI:20151637"  
 /db\_xref="FLYBASE:FBgn0000140"  
 /translation="MORTILSELLCFNPLRLGLLEVVFGBKIQMSNRDIVGLSTFI  
 LNRFRNKEQSYSKAYLTTEVAETIKKSLQKILFLLLDQAKQKRIYKHNPL  
 FVKSPEKTDILLRFSSELLANGIDITRELRIGYVLOHQRTPLDFDFVFNLLAV  
 LDRGVRTRVMEVILLRDDTLRLRVPAISRLQRIYFNVLKALGALGANFOLGDLA  
 AODIVDGHREKTLILLWOLYKFRSPKFAAATVLQKWRHRLHVLVQRIHKEML  
 RHRAATVIOAFVGHOMKVKVLFKFTERTQAAILQKFTTRYLAKOLYQSYHSIT  
 IQRWRAQLQGRHQRFRPREAREAFILQIRWRLFAKLAAAEARLQSRQQA  
 AASTIQMSRYSQIGRIQRFILRQDILFMVRMSKMSMLQKKEFQOLKRAIN  
 IQQRWAKLSMRKNADYLAISLVKYAYKATIQMRIDRNHYSLRKVVICQOR  
 LRATMKREQRNLYLRNASILVKRYRMEQMIQDRNAYLRTEKCIINVRWRAT  
 LOMRERKNVLIHOTTTKRIQIKFAKEMKQAEFLQAKVILVQKRALLQMR  
 KERQYLHREVITIKLQRRFQAQSMRPMARKYRGTOAAVSCLOHWRNHLKRRN  
 SFLOQRQAATILQRRILARNLMIKQISYAKQQAATITQTRYRAKAMQKVLYOK  
 QREAITVQRRYQNGNLEMRKQIEYQKQAVTLQKWSIRDMRLCKAGYRIELS  
 SLSTQKRWATVQARRQREIFLSTIRKRLMQAFLRCVSAKSGSSFAHCATPAHVV  
 FGVHVHLLMHGSHGSHIRGG"

## ORIGIN

Alignment Scores:  
 Pred. No.: 4.22e-287 Length: 3486  
 Score: 5267.00 Matches: 1055  
 Percent Similarity: 84.32% Conservative: 4  
 Best Local Similarity: 84.00% Mismatches: 2  
 Query Match: 55.36% Indels: 195

DB: 3 Gaps: 1

US-09-914-698-1 (1-1861) x AY094825 (1-3486)

607 ProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgLysArgSerAspArg 626

DB 1 CCATGCTCCAAAGTGGCGGTATATGTCAACAAGCAAGCTCTGCGCATCCGTAGCGATCGT 60

627 AsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeuCysPheAsn 646

DB 61 AATCTTCACCTAGACGTAGTATATGCAACGACCATCACTTGGAGCTGCTGCTTCAAT 120

647 ProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysIleGlnMetGlnSer 666

DB 121 CCCCTTTGGCTGCGCCCTTGACTGGAAAGTGGTCTTTGGCGAAGATCCAGATGCACTCT 180

667 AsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArgAsnLysCys 686

DB 181 AATCGAGACATGTAGGCCTCAGCACCTTTATCTCAATCGCTTGTTCGGAATAAGTGT 240

687 GluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTyrAlaGluThrIleLys 706

DB 241 GAGGAGCAGAGGATACAGCAAGGCATACACATCCAGGAAGATACGCGGAGACCATTAAG 300

707 LysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLys 726

DB 301 AAGCAGCTCATTTGAGAAATCTCTTTTGTGCTTTTCTCGATCAAGCTAAGCAGAG 360

727 ArgIleValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLys 746

DB 361 CGCATCTGCAAGCAATCCCTGTTGTTGTTTAAAGTCCGCACATAAAGAGACCAAG 420

747 AspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGlu 766

DB 421 GATATCTGCTGCGCTCTCTCGTGGAGCTGCTCGCCACATGTGTGATTAATACGCGGAA 480

767 LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr 786

DB 481 CTTGCTGCTGCGCTGCTGCTCTACAGCACCGCCAAACATTTTGGACGAGTTCGATAT 540

787 AlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGlu 806

DB 541 GCCTTCAACACTTGGCTGTGACCTAAGAGATGGGTGAGACTAAACCGGAGTCATGGAG 600

807 ValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeu 826

DB 601 GTAATTTTACTACCGCATGATCTAACCGCCAGTTAAGGGTCCACCATCTCTCGCCTT 660

827 GlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu 846

DB 661 CAGCGGATCTCAATGTAAAGCTAGCTCTGGGCGCACTTGGTGAAGCCCACTTCCAGCTA 720

847 GlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeu 866

DB 721 GCGCGGACATCGCGCGCCCAAGACATCGTTGACGGACATCGTGAGAAGACGCTTTCCCTG 780

867 LeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeu 886

DB 781 CTCGGCAACTTATTTTACAAATTCGCTCGCCCAAGTTTCATGCGCGCGCCACGGTGCTC 840

887 GlnLysTrpTrpArgArgHisTrpLeuHisValValIleGlnArgArgIleArgHisLys 906

DB 841 CAGAAATGTTGGCGCGCTCACTGGCTGCAACGTTGTTATCCACGCTCGCATTCGCCACAAA 900

907 GluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGln 926

DB 901 GAGCTTATCGTCGCGCACCGCGCCGCTACTGTTCATTCAGGCGCTGTTCGTCGCCACAG 960

927 MetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGln 946

DB 961 ATGAGAAAGTACGTGAAGTTGTTTAAAGCGGAACGACATCAGCGCGCAATAATCTGCA 1020

947 LysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIle 966

DB 1021 AAGTTCCACCGTCTGGTATTTGGCCCAAGACAGCTATATCAGAGCTATACAGTATTATC 1080

QY ThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVal 986

DB 1081 ACCATCCAGCGCTGGTGGCGAGCCCAACAACCTGGGAAGGAGCACCACCGCCACGCGTTGTG 1140

QY GluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLys 1006

DB 1141 GAGCTCCGAGAGCTCGATCTTTCTTCAGCGATCTGCGCGGACGACCTCTTTGCCAAA 1200

QY LysLeuLeuAlaAlaGluThrAlaArgGlnArgSerGlnLysGlnGlnAlaAla 1026

DB 1201 AAACTATTGGCGCGCGGAAACAGCCAGACTTCAGCGATCGCAAAAACACAGCGAGCT 1260

QY AlaserTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGlu 1046

DB 1261 GCTAGTTATATTAATGCAATGGCGAAGCTATCAGCTGGCAGAAATTCAGCGACAGCAG 1320

QY PheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSer 1066

DB 1321 TTCTGCGCCAGAGGAGCTCATCATCTTTGTTTCAGCGCAGGATCGAAGCAAGTGGAGT 1380

QY MetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGln 1086

DB 1381 ATGCTGAGCAGCGCAAGGAGTTCCAGCAACTAAGCGTGCGAGCTATAAATATCCAA 1440

QY ArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSer 1106

DB 1441 CGCTGGCGAGGAAGCTTTCAATGAGAAAGTGCAACGCTGATTTATTTGGCACCTCGTTC 1500

QY SerValLeuLysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsn 1126

DB 1501 AGCGTTCTTAAAGTTTCAGGCTTACAGGAAAGCCCAATCCAGATAGAGATAGATCGTAAT 1560

QY HisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMet 1146

DB 1561 CACTACTATTCCCTGCGAAAATAATGTTATCTGCTGCAACAGCGACTGAGGCGCATCATG 1620

QY LysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln 1166

DB 1621 AAAATGCGCAACAGAGGAAAATAATCTGAGGCTGCGAAATGCTTCAATAGTACTAGTCAA 1680

QY LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186

DB 1681 AAACGCTACCGCATCGCTCAACAAATGATCCAGATAGAAATGCTATTTAAGAACCCGC 1740

QY LysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeuGlnMetArgArgGluArg 1206

DB 1741 AAATGTATCATCAATGTCCAGAGCGCTGGAGAGCCACTCTGCAGATCGCTCGAAGAG 1800

QY LysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLys 1226

DB 1801 AAGAACTACCTTTCATCTCAAAACGCAACCAACGAATTCAAATCAAGTTCGCTGCCAAG 1860

QY ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVal 1246

DB 1861 CGCGAAATGAAAAAGCAAGAGCGCGAGTTCTTCAGCTGAAAAAGGTTTACCTTTGATGC 1920

QY GlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGlnTyrLeuHisLeu 1266

DB 1921 CAGAAACCTCGCGCGCTTTCGTCGAGATGCGAAGAGCGCCAGGAGTACTACATCTG 1980

QY ArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMet 1286

DB 1981 CGCGAGGTACCAATAAGCTCAGCGCAGATTCATGCTCAAAAATCAATGCGGTTCAAG 2040

QY ArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsn 1306

DB 2041 CGAGCAAGTACCGCGCACCCAGCGCTGCTGTGAGCTGCTGCAGATGATTCATGCGGTAAT 2100

QY HisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThr 1326

Db 2101 CATCTCCTTAGCAAAACGGAGAGCAACAGATTCTTCAACATGCGTCAAGCAGCAATAACA 2160

QY 1327 LeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGln 1346

Db 2161 CTGCAGGACGGTACCGAGCTCGTCTGAATATGATCAAGCAGTGTGAAGAGTTAGCCCGAG 2220

QY 1347 LeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLys 1366

Db 2221 CTGAAAACAGCAGCTATTACCATTTCAAAACCGATATAGACCAAGAGGCAATGCAAAAG 2280

QY 1367 GlnValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArg 1386

Db 2281 CAGGTGCTCTGTACCAAAAACCAAGGAGGCCATTATCAAGTGCACACGATACCGC 2340

QY 1387 GlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIle 1406

Db 2341 GGCAATCTGAGATGAGAGAGCAATTGAAGTCTTACCAAAACAGCCCGAGGAGTCAATC 2400

QY 1407 ArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArg 1426

Db 2401 CGCTTGCAAAATGGTGGCGCAGTATACGCGACATCGCGTGTGCAAAAGCGGCTACCGA 2460

QY 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArg 1446

Db 2461 AGGATTGCACTCGATTCAATTGAGCATTTCAACGCAAGTGGCGGCCACAGTGCAGGCTCGT 2520

QY 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466

Db 2521 CGCAACGGGAGATCTCTTACACCATCCGCAAGTGGGCTTATGCAAGGATTTATC 2580

QY 1467 Arg-AlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgAlaAla 1486

Db 2581 AGATGCG----- 2587

QY 1486 aValValIleGlnArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyr 1506

Db 2587 ----- 2587

QY 1506 rGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526

Db 2587 ----- 2587

QY 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnIly 1546

Db 2587 ----- 2587

QY 1546 sPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSe 1566

Db 2587 ----- 2587

QY 1566 rMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLe 1586

Db 2587 ----- 2587

QY 1586 uMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnAr 1606

Db 2587 ----- 2587

QY 1606 gTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeuAspIleAr 1626

Db 2587 ----- 2587

QY 1626 gLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLysVa 1646

Db 2587 ----- 2587

QY 1646 lGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaValIle 1666

Db 2588 -----TTAGCAGTGTCT 2598

QY 1666 uSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPh 1686

Db 2599 AAGTCGATTGGATCGTCTTTTCGGCAGCTGTGCACACCTGCTCATGTGGTGTTCGGAGTT 2658

QY 1686 eMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGlnValAspLysGI 1706

Db 2659 CATGTCACAGTTTGTCTATGCGATCATGCTCAGGCCATTCGATCAGAGGTGGTAAGCA 2718

QY 1706 nLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrVa 1726

Db 2719 GCTTATAGAGCGCTGCAGCCGGATCATCTTAATTGGCCCGCTACAAATGACACCGGT 2778

QY 1726 lAsnThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAs 1746

Db 2779 GAACACGTTCCAGAGGGCGGTTTGGTCAACATTGCCAGATGTTATTGCTGCTGTCGA 2838

QY 1746 pLysAspSerGlnIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysPr 1766

Db 2839 CAAAGACAGTGAATATTCAACACTTTGTGCACCCCTCATTTGGGTATTCGCCCACTGTCC 2898

QY 1766 oLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIleTyrMetValAr 1786

Db 2899 TAAAGCGAAGATCATTCACGACTATATGACCAACCCAGAGGCCATTACATGGTGGC 2958

QY 1786 gGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaArgLysProPr 1806

Db 2959 CGAAACTAAGAGCTCGTGGCCCGCAAGGAAAAGATGAAGCAAAATGCACGAGCGGCC 3018

QY 1806 oProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrProCysSerLeuPr 1826

Db 3019 GCCAATGACAAGTGGACGGTATTAAGAGCCAAAGATAAACTTCACGCGGTGTTCCCTGCC 3078

QY 1826 oSerLeuGluProAspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerVa 1846

Db 3079 CAGCTTGGAGCGGACTTCGGATCATCGCTACACTCCCTACAGTTCCTATCTCTGCTCCG 3138

QY 1846 lTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMetPhe 1861

Db 3139 TTACGCTTCGATACGATTTTGTGCAAGCTGCAGATCGACATGTTT 3184

RESULT 6

AF533752 9846 bp mRNA linear ROD 17-OCT-2002

LOCUS Mus musculus abnormal spindle (Asp) mRNA, complete cds.

DEFINITION AF533752

ACCESSION AF533752

VERSION AF533752.1 GI:24079963

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 9846)

AUTHORS Bond, J., Roberts, E., Mochida, G.H., Hampshire, D.J., Scott, S., Askham, J.M., Springell, K., Mahadevan, M., Crow, Y.J., Markham, A.F., Walsh, C.A. and Woods, C.G.

TITLE ASPM is a major determinant of cerebral cortical size

JOURNAL Nat. Genet. 32 (2), 316-320 (2002)

MEDLINE 22242604

PUBMED 12355089

REFERENCE 2 (bases 1 to 9846)

AUTHORS Mochida, G.H. and Walsh, C.A.

TITLE Direct Submission

JOURNAL Submitted (31-JUL-2002) Neurology, Beth Israel Deaconess Medical Center, 77 Avenue Louis Pasteur, HIM 807, Boston, MA 02115, USA

FEATURES

source

1..9846

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="Swiss Webster"

/db\_xref="taxon:10090"

1..9846

/gene="Asp"

138..9506

/gene="Asp"

/codon\_start=1

/product="abnormal spindle"

gene

CDS

```

/protein_id="AAN46088.1"
/db_xref="GI:24079964"
/translation="MATHQASCPBERGRARRPDPPEAGDPSPPVILLISHFCGVPFLCF
GVRVGRTRSLVNLHPEEPQLVLSLRAAGQGSFVAPNRCLEKPKKILITISVTW
TPLEGGVRIIVTFLVNDFLKHQAILLGNABEPPKKRSLSMTSKKIPASSKHKTRPS
KNQHNESFTISOKDRILRSPLQPCENLAMESECSPSTENKVPFPIISIRRCQSETCIP
LFTRESTAYSLNESHENTONLKVODASISOTFDNFNEVANETFINPISVCHQSGDRK
LTLAPSCSSPLNSTQTQIHLSPLDSFVNNRYTSDNLSKMNKVLSDTFPKDPAESVCL
ESQIHEVQCOTILSPDSFLNDNGLKGLKLPKSNVPLSTPQFVKDSMGHVQQTGKS
NEASQDMINEGLAIPECCOHAQTPSSSEKQNPVEVKPHKIDFTQKPKCEFDQAF
CHOSQPKHPRILSATVTRKPTNAREKLPEINPKDAKCLEGLVQORKEVSGSLRE
KRIHPSLPSTSTVARTTKKGLTSKRLSSLSRGLKKMDSSILKTPLSKTKKCR
RSIVAAQSHLTFIKPLKAALPRHPMPAAKNMFYDERWEKEKQPGQFTWMLNYLTDP
DFTVTKNVSKNAKVEIEVGLVKRDKRHLWKIOGQKQVNLWLLSINPLRLIGLST
FTSKMKVAKKVEIEVGLVKRDKRHLWKIOGQKQVNLWLLSINPLRLIGLST
VFGLIPLFNDHAKTSLRILNDPCLFCKDAEFKASKELLILAFSRDLFSGEGDLSRHLF
KLILLICFLDHAKTSLRILNDPCLFCKDAEFKASKELLILAFSRDLFSGEGDLSRHLF
LGLPVSHTQVPLDDEFPAVNLAVDLCQGVRLVTVELLITONNLSKLRIPALSRVQ
KMNVDILVOLFVRSQVPLTDEHGSATSSKDVDRHREKTLGLAKIALAFOVDISLN
LDQKEHIDFLKHTHSIKRAMSALITCPSQALTNKQDKRISGNFERVGDVQLLMDV
NAVCAFYNNKGVNFTVSDGKVLICILHHYHPCVFPDAICORTSOSVACQSGSV
LNSSESGEGCLDLSALDHSTPEMYKELLEKNHFLHLSAARDLIPAMIH
SDMNTITPDERKVTIYLSFLCARLLDLKEIRAAALITQTTWRKYLKRLDKLHQERD
AARYIQSVLNFISRRLOKNVSAALVTKCWRVSAQKRLMLKMLKLAQKNSAV
LLOAWRYSFKRFLRLKHYSVILQSEIRMKIALTSVKYLVATVITORHWRAYLSG
KRDQOIFRKLKSLIVTQPMFRWRKRLQLOTKAANTLOSAFREMHLRKOIRERSAV
VIQSWYNRHRELQIYIRSCVIVIQRRKRVLOQKQYKRRKDALILLOKHYRQK
KLHADYLOKAAATLOQAFGRKARHSYLOLGAACVLOSQYWRMQRVRFINLKK
MVILOAHIRYQLOQYKTKKAAITIQTHFRASISARRVLASQYKTRSSVYLOSA
CRGMAKAPFHALASVIKIOSYVRAVICRKTQNFKNATIKLOSIVKMQSRQYILQ
OSYFMRTRQRYLKAKCAALVIOSFYCAVEAQIOSKQNFLOVKEALICLOAAVGGCK
VRRLQKQSTAAVTIIVFRGHSQMKITQMLQSAVKLOQRYKRAVQKVAVDRIQFLKT
RAVVEVLOSAYRQVQVQOLRQHEAAVKIOSFERMAVQOQYKLLRAAAVITQOHR
AARAGKTQILAYILOHRAALVFOAAWKGMLRQIARQHOCAALIOSYRMIQORRW
SIMTKALQIOLCVRAYKQERHLYLKAAVITLOSAYRGMKVRVRAECHKAAV
TIQSKFRAYRTQKVTYRTSAIVTORWVNIKITTOOHOYLNLRAAVOVQAAVYRG
LIVRRRTIOHMAATLIEAMFMRQSVRVYLMKMTAALIOVRYEAVYLGKIOHEKYL
ATLKAIFLOAGVGRARRVTRVMKFAATLLOSHFRGHRQOTYFHLRKAATVQQR
YRAVKEGSAEIVQYRSLRSLVLLIOAFRGLRTRHLKAMHLAATLLORRPRITFAMR
KFLSLRKTAIMIQYQYRRLYAKYSROQLLEKAVIKIOSYRGMVVRKRVKQHRRA
VITQATFMHGAQVRYOHLKRAVVIQVHTAAELQORKAAVILQAAVYRGMKTRSHLX
TMHSSATLIQSOFAFRIVRRFIALRKAALFQVRFRATLVAKHKLHOFLOKRAAIT
IOSYRRLMWOKKIQEMHBAALIOAFEMHRTVVAFHWKASIRIOCCYRTVTK
LOKEKLIKESQHSAAVILQSTYRMTQRCFTQORWAKVITQYTRANKRQDILLYCA
KEETPLLOMHPQGLNTAKQOQHGGAAMITQKHFRAFKAARLMEARFGQACRKYK
KYLKSVBAACRIQAWYRMRRAHKYLLTAKVNIIEGLYSAQLARRPFLKGRAAAI
IORKWARTSVIRGARENLRHRAACVIOAHFRGQARQSFLOQRSAAVILIOHVRAMV
AAKOBRIKVIKLKSTVVVOALVRGMLVRKVSEOKAKTRLFPHFTAAHACWALIKIO
PAYERLHVTLNAKHMDSVIFIORWFRLOKRFLEOYHKILSTPRAHACWLOQDR
AASVIOKARFPLICRREKITSCTRIOALWRGYSWRKNDHTEIKAIRSLRAVST
TVEENKILRTERALHLLITKHLSDALDHLVTVTRSLSPGCCENNAESGAVSTI
FVWIRSNRSPCEMVGVAVQVLLNAIKDYTKTAAVYEAENCVDITLLELLQVREKP
GDRVAEKASAFITRCCILLVLLKTEQAFDAQSKSVTDRIYRLKYFTVPKHKVNTQ
GLEFDKQKNSCVGPECIPIERTMTKRLVLSRLKPQWVLRDRDNVEEITNSLOAIQLVMDTL
GISY"

```

ORIGIN

Alignment Scores:					
Pred. No.:	2,4e-67	Length:	9846		
Score:	1374.00	Matches:	552		
Percent Similarity:	38.11%	Conservative:	371		
Best Local Similarity:	22.7%	Mismatches:	741		
Query Match:	14.4%	Indels:	758		
DB:	10	Gaps:	77		

US-09-914-698-1 (1-1861) x AF533752 (1-9846)

QY	1	MetGluLeuValTrpSerProValLeuGluAlaCysLysGluThrLeuGlnLeu	20
DB	429	ATTTCGGTTACCTGGAGCCACTGCGAGAAAGGGGAGTGGGGAGATTGTACATTTCTT	488

QY	21	AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro	40
DB	489	GTAAT--GATTTCTGAAAGACCAGGCTATATTA--GGAAATGCGAAGAGCCT	542
QY	41	VallysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr	60
DB	543	AAGAAGAAAAAGAGAACTCTTTGGAATACCAAGTAA--AAGATTCCAGGCC	590
QY	61	GlyAlaGlyLysThrMetLysSerVal--ValSerAlaAlaVal	74
DB	591	TCITCAAAACATCAAAAAGGACTTCCAAAACCAACATTTTAATGATCATTTACTATT	650
QY	75	GlnGlnLysLysArg--MetSerAla	82
DB	651	TCACAAAAAGACAGATTTAGGAGCCACTGCGAGCTTGTGAAAAATCTGGCTATGAGTGA	710
QY	83	AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSer--ArgProAla	101
DB	711	TGCTCTTTCCCAACAGAAAAC--AAAGTCCCCACCCCATCCATTTAGTCTTATT	761
QY	102	AlaTrpAlaHisProProGlnAlaProLeuVal--GluLysAsnValTrpLys	119
DB	762	AGAGAAAGCCAGAGTGAACCTTGCTTGCACCTGTTTACCGGAGTCCACTGCCTATTCA	821
QY	120	ThrProGlnGluProValTrpLysSerProGlnProArgSerLeuLys--GluAsn	138
DB	822	TCTCTTCATGAA--TCTGAAAAATACACAAAATTTAAAAAGTACAAGAT	866
QY	139	LeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuAspPhe--Thr	157
DB	867	GCCAGCATTTCCAAAACCTTTGATTTAATGAGGAAGTCGCAATGAAACTTTTATTAAT	926
QY	158	ProLeuThr--GluThrArgGlyLysGlyGlnAlaThrIlePheProAsnAsn	174
DB	927	CCATAGTCTGTCTCACCAGAGTGAAGGATAGGAACTCAGCTTGGCCCCAAAC--	983
QY	175	LeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAsp--	193
DB	984	--TGTTCTTCCCTTTGAAATAGTACACAGACCACCAATACACTTTTCTAAGTCCAGATCT	1040
QY	194	--MetArgProArgGlyIleThrProAspAspLeuGluAsp--GlnProAlaThrAsn	211
DB	1041	TTGTGAAACACACAGGTATACGCTGTGATACCTGATACCTGGAATCAATGAAAAATGTTTATCA	1100
QY	212	LysThrPheAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp--	230
DB	1101	GACACATTTAGGAAGATCCCCGAGAGTCTGTATCTCTGGAATCCCAAACTGTACATGAA	1160
QY	231	--CysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrThrIleVal	249
DB	1161	GTTTGCCAAACAATC--TTAAGCCAGATCTTTTTTAAATGATAATATTGAGACTTAAA	1217
QY	250	HisAlaThrHisThrArgAlaAlaCysIleHisGluGluGluGlyProSerProPro	269
DB	1218	AAAGGGCTAAACTTCAATCCGTC--AATCCCGTT	1250
QY	270	ArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySer	289
DB	1251	TTGTGCGCCACACAAATTT--GTAAAGGATAGCATGGGACACGTAGGTGAG	1298
QY	290	ProLeuArgLysTyrSerGluSerMetLysLysLeuSerLeu--Leu	304
DB	1299	CAAAAGGGTAAAGTCAATGAAGCTTCTCAGGATTTGGAGATAAATGAAGTTTGGCATAT	1358
QY	305	SerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIle	324
DB	1359	ACCCCTGAGTGTACGATGCA--CMAACTCCAGCAGCAGATCTGAAAAACAA	1409
QY	325	ArgSerIleGlu--	328
DB	1410	RACCCCTGTAGAGTGAAGCCACATAGTATGATTTTACAAAGCAAAAGCCCAAGACTGT	1469
QY	329	-----GlnAsnArgTyrTyrGlnGluGlnGlnIleGlnIleLysAlaLysAspLeuAsn	346





QY 876 -----SerProLysPhe 879  
Db 3633 GATCAGAAAGTACTCCAGAAATGTACAAAGAGCTTCTAGAAAACGAGAAAGAAATTTTC 3692  
QY 880 His----- 880  
Db 3693 CATTTGGTGGATCTGCAGCTAGGATCTGGAGGATACCTGCTATGATTTCATTCATCA 3752  
QY 880 ----- 880  
Db 3753 GACATGTCAAATACAATTCCAGAGAAAAGGTGTTATCACCTATTTGTCTATTCTTTGT 3812  
QY 881 -----AlaAlaAlaThrValLeuGlnLysTyrTyr 890  
Db 3813 GCAAGGCTTTGGATCTCCGTAAGAAATAGAGCTGCTGCAGCTTATACAGACAACTGG 3872  
QY 891 ArgArgHisTyrLeuHisValValIleGlnArgIleArgHis----- 905  
Db 3873 AGAAATATAACTA-----AAAGAGATCTGAACACCATCAGGAAGAGAC 3920  
QY 906 -----LysGluLeuMet 909  
Db 3921 AAAGCTGTAGATTATTAGTCACTGTTGTTCTCAATTTTCTAAGTAGACGAAGATTGCAA 3980  
QY 910 ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys 929  
Db 3981 AAAAACGTCAGTCAGCTCTGTCATTCAGAAATGTGGCGAAGAGTCTCAGCCCAAGG 4040  
QY 930 TyrValLysLeuPheLysThrGluArg-----ThrGlnAlaAlaIleIle 944  
Db 4041 AAGTTGGGATGCTGAAATGAAATGCAAACTGCGCAAACTTCAAAATAAATCAGCAGTGTCTC 4100  
QY 945 LeuGlnLysPheThrArgTyrLeuAlaGlnLysLeuTyrGlnSerTyrHis--- 963  
Db 4101 ATTCAGGCTTATTGGAGAAGTACTCCACCAGAAAACGGTTTCTCAGGTTGAAGCACTAT 4160  
QY 963 ----- 963  
Db 4161 TCAGTCACTCTCAGTCTAGGATAGAAATGCTTCTACATCTTACAAAGCATAT 4220  
QY 964 -----SerIleIleThrIleGlnArgTyrPheArgAlaGlnGlnLeuGlyArgGlnHis 981  
Db 4221 CTTTGGGCCACAGTTACAAATTCAGAGGCAATGGCGTGCCTATTTAAAGTGGAAACAGAT 4280  
QY 982 ArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGln-----ArgIle 998  
Db 4281 CAGCAAAATATTTAGGAAGCTGAATCTCTCATCCCTTGTAAATCCAGTTTATGTTTCCAGACGG 4340  
QY 999 TrpArgArgGluLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGln 1018  
Db 4341 TGGAAAGCGACGCAAGCTGAGTTACAAACCAAGCTGCAGTAACATTCACGCGAGCTTTT 4400  
QY 1019 ArgSer-----GlnLysGln----- 1023  
Db 4401 AGAGAATGGCATCTCAGGAACAAATACGAGAAAGATCCGCTGTTGTGATACAGTCAATGG 4460  
QY 1023 ----- 1023  
Db 4461 TACAGAATGCACAGGGAATTCAGAAATACATTTATATTAGCTCTTGTGTTATTGTCATC 4520  
QY 1024 -----GlnAlaAlaSerTyr----- 1029  
Db 4521 CAGAGAAAGTTCGTGCTTTCAGCCCAAGATTATTAAGAAAGGAAGATGCCATC 4580  
QY 1030 -----IleGlnMetGlnTyrArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPhe 1047  
Db 4581 CTGACTCTCCAGAAACATTTACAGGCGCAGTCAGAAAGGGAAGCTAGCACAGCTGACTAT 4640  
QY 1048 LeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArg----- 1062  
Db 4641 TTGCAGAAACGAGCTGCCACCATTCGACTGCGAGGCTGCATTCAGGGGATGAAGGCTCGC 4700

QY 1063 -----SerLysTyrSerMetLeu 1068  
Db 4701 CATTCATACAGGTTGCAGATTGCAGCTGCCTGTGTTCTTCAGTCTCTATTGGAGATGAGA 4760  
QY 1069 GluGlnArg----- 1071  
Db 4761 CAGGAAGAGTTAGATTTTAAACCTTAAAGATGGTTATTAAAGTTGCAAGCACACATA 4820  
QY 1072 LysGluPheGlnGlnLeu-----LysArgAlaAlaIleAsnIleGln 1085  
Db 4821 AGAATATACCAACACATACAGAAATATAGAAAGTAAGAAAGCAGCATTACAAATTCAG 4880  
QY 1086 GlnArgTyrPheArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArg 1105  
Db 4881 ACTCAITTTCCGAGCTTCCATTTCCAGCCAGGAGGTTCTAGCATCTTATCAGAAACCAAG 4940  
QY 1106 SerSerValLeuLysValGln-----AlaTyrArg 1115  
Db 4941 TCTTCAGTCATGTTCTACAGCTCTGCATGTAGAGGGATGCAAGCAGAAAGCAATTTCCG 5000  
QY 1116 LysAlaThrIleGlnMetArgIleAspArgAsnHisTyr-----TyrSerLeuArgLys 1133  
Db 5001 CATGCTTTAGCATCCGTTATAAGATTCAGTCATATATTCGGGCTTATATATGTAGGAAG 5060  
QY 1134 -----AsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGlu 1150  
Db 5061 ACATTTCAGAACTTCAAAAATGCCACAATAAAGCTGCAGTCAATTTGTCAAAATGAAACAA 5120  
QY 1151 GlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArg 1170  
Db 5121 TCAGGGAAGCAGTATTTACAAATTAAGAGCAGCTGCATCTGTTTATCCAGCGTGTACCGT 5180  
QY 1171 MetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle 1190  
Db 5181 TCCCAAAATCGGCTCACAAGAGAGAAAGAGTACATACAGGTGCGTGAATCTGTATC 5240  
QY 1191 AsnValGlnArgArgTyrPheAlaThrLeu----- 1200  
Db 5241 AAATGCACTCACACTTTAGAGGATGCTGCTCAGAAAGCAGTTGAGGTTGCAATGTAAA 5300  
QY 1201 -----GlnMetArgArgGluArgLysAsnTyrLeuHis 1211  
Db 5301 GCTCCCAATTTCTACAGTCTTATTCAGAAATAGAACACAGGAGCGCGCTACTCTGAAG 5360  
QY 1212 LeuGlnThrThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLys 1231  
Db 5361 ATGTGCAAGCGGCTCTGTTCATTCAGAGTTCTACTCTGCTACAGAGCACAAATCAGT 5420  
QY 1232 GlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValValGln----- 1247  
Db 5421 CAAAGGAAGAACTTCTTGCAGGTCAAAGAGCAGCTATTTGTTGCGAGCGAGCTTACAGA 5480  
QY 1248 -----LysArgArgArgAlaLeuLeuGlnMet----- 1256  
Db 5481 GGCTGCAAGTACCGCGCAATCAAGCAGCAGTCTACGCTGCTGTGTACAAATTCAGAGG 5540  
QY 1257 -----ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLys 1272  
Db 5541 GTTTTGTAGGCCACAGTCAGAGGATGAAATACCAGACTATGCTTCAGTCTGCTGTGAAG 5600  
QY 1273 LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly 1292  
Db 5601 ATTCAGAGATGTTGAGGCGCACAGAAAGGTTTCCTTATGATGAGAAATTCAAATTTCTAAAG 5660  
QY 1293 ThrGlnAlaAlaValSerCysLeuGlnMetHisTyrArgAsn----- 1306  
Db 5661 ACTAGAGGCTGTGGTTTGCTCCAGTCTGCTTCCCTTACCGTGGCTGGCAGGTTGGGCAACAG 5720  
QY 1306 ----- 1306  
Db 5721 CTCAGAGGCAACATGAGGCTGCAGTGAAGATCCAGTCTACGTTTGAAGATGGCCGTGGCC 5780  
QY 1307 -----HisLeuLeuArg----- 1310





```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
1..10434
/notes="ASP"
/codon_start=1
/product="abnormal spindles"
/protein_id="AAN0011.1"
/db_xref="GI:23884551"
/translation="MANRRVGRGCVSEVPTERRPPAGLRGAEEERASSPPVLSLSPH
CRSPFLCGVLLGGRREISLALDNPNREVAEKISHFPAADLGSFQSRFCVLPKPE
KIVISVNTPLKGRVREIMTFLVNDLKEQAIIILGNAPQKKKRSRLLTKKKIS
ASTSHSRVSNQVNTFVSQVDRVSPLOACENLANNEGPPENNLSLILEENK
IPISPIFAPNEHCAGCLPLSVPRSTYSLSHASENRELLNVHSANVSNKEKAV
TETSPVNVNQGENSKSLTPNCSTLNITQSHFLSPDSFVNNHGANNEL
VTCLSDFMFKDNGQPHLESTIAHEYOKILSPDSFIKONJLQNGLESSESNPILS
PQLKDMNMYMTSQDTQKVLPSNENSQPOSEDWRKGEVSPRIPECCQSGSKAI
FEELVEMSNYSYFIKONPKFSVQDISSSHNKQPKRPILSATYTKRATCTREN
QTEINKPKAKRCLNSAVGEHEKILNNOKEDEPHSLPIIDPILSKSKYKNEVTPES
TITASVARKKSDGSMEDANVRVAITEHEVREIKRIHFSPSEPKTSVKKTKNVTPI
SKRLSNREKLNKKKTDLSIFRTPISIKTNKRTKPIILVAQSSLTPIPLKTDIPRHM
PFAAKNFYDERWKEKQEGFTWMLNFIPLPDDFTVKINISVYNAAILHLEIENQHKI
SVRPAPTEKEMSELYATARCENLNRRAACRLFTSEKWKAKIKLEIEIERRLLVRK
DRHLKDVGGERKVLNLLSYNPLMLRIGLETTYGELISLEDNSDVTGLAMFILNRL
WNPDIAAEYRHTVPHLYRDGHEALSKFLKLLLVCFDYAKISRLIDHDPCLFC
KDAEFKASKEITLAFSDPFLSGEGDLRHLGLLGLPVNVHVTQPFDFVAVNLAVFL
OCGVLRVWTELLTQWDLKSLRIPALSRLOKMHNDIVLOVLKSGIELSDHGNT
ILSKDIDVRHREKRLRLKIAFAFQVDISLNDQLBEAFLKHTKSIKTSISLLSC
HSDLLINKKQSGSFQSENIKULMDVNAVCAFYNNKKNVENFTISDGRVLCT
LIHYHPYVFPDAICQRTQTVTECTQSVLNSSESDDSLDMSLKAFDHEBNTSE
LYKELLEKNKFLHRSVARDLGGIPAMINHSMSNTIPEKVVITYLFLCARLID
LRKIRARLIQTTWRYKLTDLRHOEREKAARIITQLAVINFLAKORLKRNVNLA
VIQKXWVLAQRLLKLEKLEKQVQKASLIQGYRBYSTOFRLKLYSILIQ
SRITKTYKRYLWATVITQHRNAYLRKQDQOQRYEMLSKSTLLIQSMFRKWKQ
RKMSQVATQILQRAVREMLRKOAKEENSAIIQSWYRNHKEIKIYIRSCVVI
QKRPRCAQVILKRRRESILTIQKYKAYLKGKIERNTYLOKEAAAIQLOAFRRIL
AHNLRCROIARAACTVQSWYRQDRVFLNKKTIIFQOAHVRKHQKQYKMKKAA
IITQHFRAVIFAMKVASYOKTRSAVILQSAVGMQARKMYIHILTSVTKISQYRA
YVSKKEFLSKNATIKLOSTVMQKQVXLIHPAALFLOQCVRSKKIAQKREYVM
QMRSCIKQFVGYLVRKQMLQKAVISLOQYFMRKQAYILKMYKALIVIQNY
YHAYKAQNRQNLQVKAATCLOAAYGYKVLQKQSLAALKIQASFRGVVRVRLTKMQ
KYQSLQSIKIQRYRAYTLHTDTRFLKTKAAVILQSAVGMQVRLKQIRREHOA
ALKTQSAFMAKQKQFLKTAALVIFQNFRAWTAGKQCMYELHVLAVLQSWM
KGKTLRLOQHQKCAIIIOQSVYRMVQKQKWKMLKKAALLIOKYRAYSTGRONHL
VLKTAVALVTLOSAVGMQVVKRIKDCNKAATVIOQKYRAVTKKXYATRASALIIQ
RWYRGKITNQHKEYLNLKTAIKIQSVYRGIVRHHIOMHRAAFIKAMFKMHOS
RISYHTMRKAAIVIOVRAYYQGMQREKYLILKAVKVLQASFRGVVRVRLTKMQ
TAATLIQSNTRYRQTYFNKRIKTKTQOQRYWAMKERNITQFORYNKLRHSVYIOA
IFRGKARRHLKMHIAATLIQRRFRITLMRRERFLSKKTAIILQRYKRAHLCTKHL
QFLQVONAVIKIQSVYRMVLRKMRHRAATFIQSTFRMHLRMYRALKQASVVI
QOQYQANPAKQHQVILROPHSAVILQAFEGMKTRHLSKMHSSATLIQSRFSRL
VRRPFLSKKATIFQRYKRAVICAKHLYOFLHRAATIIQSSYRRLMVKKLQEM
QRAAVLIQATFMHRTYITFTQWHSALIIQOHTYTRAAKLORENTIROMHSVVIQ
AAYFGMKQRLKREKASIVIOQYRYMYQYFYQKLOKAWATKIIOEKYRANKKKQV
FOHNLKKTQVQAGQDMNIIKQIOHQAAIIIOKHCKAFKIRKHYLHLRAWTSVI
TAYVILQNYELAVVITKQSVYRGVVKDIQNMHRAATLIQSFYMRHAKVDYETK
TAIVILQNYELAVVITKQSVYRGVVKDIQAFEGMKVROKLKNSVEKVAIV
NQSALCCVRSQTOHVAQSEGVIQEWYKASGLAQSBQAEVHSQRAAVTIQKAFCM
VTRKLETKAALRIQFLOVAVYRRVFRVQKRAATIIQHYFRFTQKQFLLYRKAA
VVLQNYRAFASLKHQVQVLIQSRVSIIOQASKGFQKQKFOEIKNSTIKIQAMR
RYRAKYLKVKAKKAIQAWYRCWAHEKYLAILKAVKIIOGCFYTKLERFTFLNRA
SAIITQKRWAILPAKIAHEHFLMKHRAACLIQAHVYRGVQKQVRLQKSAALIIQ
KYIARBAKHERIKYIEFKSTVILQALVRGLVRKFLQRAKIRLHPTAAAYH
LNAYRIQAYKILAVNANKQVNSVICIQWFEARLOEKEPIOKYHSIKKEHGOE
CLSQNRNRAIVQAVRFLKQKQKFTSGIILKIQLWNRGYSWRKNDCTKIKAIL
SLQVNRREIRENKLKPTALHLYLTYKHLAISALEALHLEVVTVLSPCCENMAQ
SGAISKIPVLRSCNRSYKPCMEVIRYAVQVLLNSYEKTTSAVYDVENCNLIILEL
QIYREKQNVKADGGSIPTKTCILLAILKLTNRASVRSRQVDRIVSLYKLTAH
KHKNTERILYKQKNSSISPTFIPETPVTRIVSRILKPDVLRDNNMEETNPLQAI
QWMDTLGIPY"
```

Alignment Scores:		1.6e-66	1360.00	Length:	10434
Pred. No.:		1360.00	Matches:	478	
Percent Similarity:		39.81%	Conservative:	362	
Best Local Similarity:		22.65%	Mismatches:	630	
Query Match:		14.29%	Indels:	640	
DB:		9	Gaps:	67	
US-09-914-698-1 (1-1861) x AF509326 (1-10434)					
Qy	5	TrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuIleAspAsnArgAsn	24		
Db	328	TGGACACCATCTCAAGAGAGCCGAGTAGAGATATGACATTTCTTTGTAAT--GAT	384		
Qy	25	PheArgLysGluValMetIleLeuLysSerLysSerAsnGlnProValLysAsnPro	44		
Db	385	GTTCGTGAACACCAAGCTATATCTAGGAAATGCAGAGACAGAAAGAAAAGAGG	444		
Qy	45	ArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThrGlyAlaGlyLys	64		
Db	445	AGTCTTTGGGATACCA--	462		
Qy	65	ThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMetSer	81		
Db	463	---AAAAGAGAGAAATTTTCCAGCTCTACAGTCACACAGAGGGTTCAATATTCAG	519		
Qy	82	-----AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArg	99		
Db	520	AATGTTAATAAAACATTTAGTGTGTTTCCAAAAAGTTGACAGAGTAGGAGCCCACTCAA	579		
Qy	100	ProAlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLys	119		
Db	580	GCTTGTGAAACTTGGCTATGATGAATGAAGGGCGGTCCCCCAACAGAAACAATCTTTA	636		
Qy	120	ThrProGlnGluProValTyrIleSerPro	130		
Db	637	ATACTTGAAGAAATAAATACCCATACCTATACCTTACCTGCTTCAATGAATGCCAT	696		
Qy	131	-----GlnProArgSerLeuLys	141		
Db	697	GGTCAACTGCTTGGCCACTCTCTGTACGTCGATCTACTACTCATCTCTCTTCATGCA	756		
Qy	142	MetThrProGlyAsnLeuLeuAspVal	157		
Db	757	TCAGAAATAAGGAACTATTAAATGTACACAGTGCCACAGTTCCTTTCAAAAGTTCTTTAAT	816		
Qy	158	-----ProLeuThrGluThr	165		
Db	817	GAGAAAGCTGTAACCTGAACTTCCTTTAATCTGTAATGTTAATGGCCAAAGAGGAGAG	876		
Qy	166	GlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThrProLeuLysGly	185		
Db	877	AATAGTAACCTT-----	900		
Qy	186	AsnValLysSerCysAlaAsn-----AspMetArgProArgArgIleThrProAspAsp	203		
Db	901	AACCTGTTCTTCAACTTTGAAACATTTACACAAAGCCAAATACATTTTCTTAAGTCCAGATTCT	960		
Qy	204	LeuGluAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsn	223		
Db	961	TTT-----GTAATAATAGTCATGGAGCTAAT	987		
Qy	224	IleSerLeuAspThrLeuAspCysSerArgIleAsp-----GlyGlnProHisThr	240		
Db	988	AATGACTAGAAATTAGTAACATGCTTTTCATCAGATATGTTTATGAAAGATAATTACAG	1047		
Qy	241	ProLeuAsnLysThrThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle	260		
Db	1048	CCTGTGATTTGGAAATCAACAATTTGCACATGAAATTTATCGAAATTTTA-----	1098		
Qy	261	HisGluGluGluGlyProSerProArgThrProThrLysSerAlaIleHisAsp---	279		
Db	1099	-----AGTCCA-----GATTCTTTCATAAAGATAAT	1125		

280 -----LeuLysArgAspIleLysLeuValGly---SerProLeuArgLysTyrSerGlu 296  
1126 TATGGACTAATCAGGATCTAGATCAGATCAGTCAATCTATTTATCCCTAATCAA 1185  
297 SerMetLysAspLeuSerLeuSerProGlnThrLysTyrAlaIleGlnGlySerMet 316  
1186 TTTTAAAGATAAATGCGCATATATGTGTACATCTCAGCAAAACATGTAAAGTACCAATTA 1245  
317 ProAsnLeuAsnGluMetLysIleArgSerIleGlnAsnArgTyrTyrGlnGluGln 336  
1246 TCAAAATGAAATTTCTCAAGTCCACAGTCTCTCTCAAGAT-----TGGAGAAAAGT 1296  
337 GlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAla 356  
1297 GAAGTTTCGCCCGTATCTCTGAATGTCAAGGTTCAAATCTCCAAAGCTATTTTGAA 1356  
357 GlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSer---ArgPhe 375  
1357 GAACCTAGTAGAAATGAAGTCAAAATTAATCTACAGATTTTATATAAAACAAATAATCTTAAATTT 1416  
376 Asn---LeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnPro--- 393  
1417 TCTGCGATTCAGGATATTTCTAGTCTAGCCACAATAAACCACTAAGAGACGCTCCAATA 1476  
394 -----HisLysArgArgSer----- 398  
1477 CTTTCTGCCACTGTTACTAAAGAGAGGCCCTGTACACAGAAAACCAAACTGAGATT 1536  
399 -----HisGluLeuSerPhe 403  
1537 AATAAACCAAAAGCAAAAGATGCTCAACAGTGCAGTGGTCAACATGAAGAAAGTAATA 1596  
404 SerAspAlaProSerAsnGluSerLeu----- 412  
1597 AATAATCAAAAGAAAGAAAGATTTTCAATCTTATCTTCAATATATAGATCAAAATATTA 1656  
413 -----TyrArgAsnGlu-----ThrValAlaIleSer 421  
1657 AGTAATCTAAGAGTTATATAAACGAGGTAAACCCCTCTCGACAACAGCTTCAGTTGCT 1716  
422 ProProLysLysGln-----ArgValGluAspThrThrLeuProArgSerAlaAlaPro 439  
1717 CGAAAAGAAAGAGCGATGGAAGCATGAAGATGAAGATGCAATGTG---AGAGTTGCAATTACA 1773  
440 AlaAsnAlaSerAlaArgSer---SerSerAlaHisAlaTrpProHisAlaGln----- 456  
1774 GAACATACAGAGTGCAGAGAAATCAAAAGAAATCCATTTTCTCCCTCAGAGCCCTAAACAA 1833  
457 -----SerLysLysPheLysLeuAla 463  
1834 TCAGCTGTTAAGAAAACAAAATGTGCAACACCCATCTCAAAACGTATTAGCAACAGA 1893  
464 GlnThrMetSerLeuMetLysLysPro-----AlaThrProArgLysVal 478  
1894 GAGAAATTAACCTGAAGAGAAACCTGATTTATCAATATTCAGAACTCCAATTTCTATAA 1953  
479 ArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyr----- 495  
1954 ACAACAAAGGACAAAACCCATATTCGTGTGGCACAGTCCAGTTGTGACCTTTCATATAAA 2013  
496 ---MetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThrIleAspProPhe 514  
2014 CCATTAACACAGATATTCACAGACCCGATG-----CCATTT 2052  
515 LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLys 534  
2053 GCTGCAAAAAACATGTTTATGATGAACGCTGGAAGGAAAAGCAGGAACAGGCGCTTCACT 2112  
535 LysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsn---Asn 553  
2113 TGTGTTAAATTTTATATAAC---CCTGATGACTTCACTGTATAAAAAACAATATTTCT 2169

554 LysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValVal----- 571  
2170 GAAGTAAATGCTGCTACTCTTTTGGGAATAGAGAAATCAACATAAAATAAGTGTTCCT 2229  
572 ---AlaProThrLysGluGlnSerMetAsn---TyrLeuThrLysTyrArgLeuGlu 589  
2230 AGAGCACTACAAAAGAGGAATGTCTCTCAGAGCTATATCTCGTCCGTAGGTAAAC 2289  
590 ThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSer 609  
2290 AGACTAGTGTGTCAGCATGCGCTTTGTTACTTCTGAAAAAATGGTTAAAGCTATTAAA 2349  
610 LysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHis 629  
2350 AACCTTGAATTTGAATTTGAAGCTAGCGGTAAATTTTTCGAAAAAGATAGACACCTATGG 2409  
630 LeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeuCysPheAsnProLeuTrp 649  
2410 AAAGATGTGGGAACGCTCAGAAAGTCTGAATTTGGCTGTGTCTTACATCTCTTTGTGG 2469  
650 LeuArgLeuGlyLeuGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAsp 669  
2470 CTTTGAATTTGGTCTAGAGACAACCTTATGGAGAACTCATATCTTTTGGAAAGATAACAGTAT 2529  
670 IleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArgAsn-----LysCys 686  
2530 GTACAGGGTTGGCTATTTTATCTGAATCGCTTCTTTTGGAAATCTCTGATATAGACGT 2589  
687 GluGluGlnArgTyrSerLysAlaTyrLeuThrGluGluTyrAlaGluThrIleLys 706  
2590 GAGTATAGACACCCACTGTTCTCTCACTGTATAGAGATGCTCATGAAGAAGCTTTGTCC 2649  
707 LysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLys 726  
2650 AAGTTTACATTTGAAAGTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2709  
727 ArgIleValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLys 746  
2710 AGACTCATTTGATCATGCTTGTCTCTCTGTAAAGATGCCGAATTCAGGCTAGTAA 2769  
747 AspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGlu 766  
2770 GAAATCCTTTTGGCTTTTTCAGGAGATTTCTTAAGTGTGAAGTGCACCTTTCCCTCAC 2829  
767 LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr 786  
2830 CTTGGCTTATCGGATTAACCTGTTTACCATGTTTCAGACACCATTTTGAATTTGATTTT 2889  
787 AlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGlu 806  
2890 GCGGTTACAATCTTCCCGTAGACTTGCAATGTGGAGTGGCCCTTTGTGGACCATGGAA 2949  
807 ValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeu 826  
2950 CTTCTCACACAGAACTGGGACCTCTCAAGAAACTCAGGATTCGGCAATAAGTCTCCT 3009  
827 GlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu 846  
3010 CAAAAGATGCACATGTTGACATTTGTTCTTCAAGTTCTTTAAATTCAGGAAATGAATTA 3069  
847 -----GlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLys 862  
3070 AGTGATGACGATGGAATACAAATCTATCTAAGGATTTTGTGGATAGGCACAGAGAAAA 3129  
863 ThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArg----- 875  
3130 ACTCTCAGTTGCTTTGGAAAAATAGCGTTTGTCTTTTCAGTGGATATTTCCCTTAACCTTA 3189  
875 ----- 875  
3190 GATCAATTAAGGAAGAAATTCCTTTCTTAAACACACAAAGAGTATAAGAAAAACAATA 3249  
875 ----- 875





```

CDS
/ds_xref="MIM:605481"
127..3555
/codon_start=1
/product="ASP protein"
/protein_id="AAH34607.1"
/db_xref="GI:21961301"
/db_xref="LocusID:259266"
/translation="MSTRAYTARCLNRLRRAACRLFTSEKMKVAKIKKLEIEIEBRL
IVKDRHLKDWGRQKVLNLLSYNPLMLRIGLETYYGELISLEDSNDVTLGLAMFIL
NRLWNPDIAAEYRHPYVPHLYRDGHEEALSKFTLKKLLLLVCFDLTAKIIRLIDHP
CLFCKDAEFASKEIILAFSDLSGSDLSRHGLGLGVNHWQTFDFBFAVTNL
AVDQCQGVRIWVWMLLQWDLKSLRIIPAIISLQNMHNDVIIVLQKSGEISDE
HGNTLISKDIIVDRHREKTLRLMLFAFQVDIISLNDQLKEELAFKHTKSIKKTIS
LLSCHEDLNLKKGKSDSPFOYSENIKLMDVNAVCAFYNNKVFNTVSPSDGR
VLCYLIIHYHPFPDAICORTQTVECTGTGSVAINSSSESDDSLDMSLKAFDHE
NLSLYKELLENEKNNHVRSAVDRLDGGIPAMINHSNNTIIPDEKVIYIYLSFLCA
RLDLREKEIRRAURLIQTKRYKLTDLKRHOEKEAARIIOIAVINFLAKORURKV
NAALVIQKYVRVLAQRLKMLKEKLEKQNKAAASIOAMWRYRAKYLCKVKAAC
KIOWYRCWMAHKEYLATLKAVKIIQGCFTYKLBRTPLNVRASAIIIQKRWAILPA
KIAHEFLMKRHEAACLIQAHVRYGKGRQVFLRKSAALIIOKVIRARBAKHRIK
YIERKSTVILOALVRGWLVRKLEQRAKIRLLHFTAAAYHNLNARIORAYKLYLA
VKNANKQVNSICIQMPARLOKRFIOKTHSEKLEHEGQECLOSNRRAASVIOXA
VRFELKROKFTSGGIKIQALWRGYSWRKNDCTKIKATRLSLQVNRREBENKL
YKFTALALHYLLTYKLSAILLEALKHLEVTVLSPLCENNAQSGAKIPIVLIRCN
RSIPCEMIVRYAVQVLLNVSKEKTSKAVDVENCIDILLELQIYREKPKGVADKG
GSIPTKCLLAILLKTNRASDVRSKRVVDRIYSYKLTFAHKHKNTERILYKQKK
NSSISIPPIETPTVTRIVSRLEKPDVLRDMMEEIINPLQAIQWMDTLGIPY"
1216..1395
/misc_feature
/note="CH; Region: Calponin homology domain"
/db_xref="CDD:smart00033"
2155..2769
/misc_feature
/note="COG5022; Region: Myosin heavy chain [cytoskeleton]"
/db_xref="CDD:COG5022"

ORIGIN
Alignment Scores:
Pred. No.: 4,746-43 Length: 3760
Score: 932.50 Matches: 367
Percent Similarity: 40.63% Conservative: 229
Best Local Similarity: 25.02% Mismatches: 477
Query Match: 9.80% Indels: 395
DB: 9 Gaps: 51

US-09-914-698-1 (1-1861) x BC034607 (1-3760)
Qy 500 IleAsnProAspPropheAlaalaThrThrThrIleAspPropheLeuAlaSerThrMet 519
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7 TTAACCCCTGATGACTTCACTGTAAAAACAAATATTTCTGAAGTAAATGCTGCTACTCTT 66
Qy 520 TyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysThrLeuAsnAla 539
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 CTTTGTGGAAATAGAG-----AATCAACATATA----- 93
Qy 540 LeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLys 559
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 ---ATAAGTGTCTCTAGA----- 108
Qy 560 LeuPheAsnGluValArgAsnLysGluLeuValAlaProThrLysGluGlnSer 579
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 -----GCACCTTACAAAAGAGAAATGTCT 132
Qy 580 MetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu 598
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 CTCAGAGCTTATCTCGGTGTAGTTAAACAGACTACGTCTGCGAGCATCGCGTGTG 192
Qy 599 PhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGln 618
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 TTTACTTCTGAAAAAATGGTTAAAGCTATTAAAAAGCTTGAATTCGAATTCGAAGCTAGG 252
Qy 619 AlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIle 638
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 CGGTTAATGTTCCGAAAGATAGACACCTTATGGAAGAGATGTGGGAGAACGTCAGAAAGTC 312
Qy

```

```

639 LeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPhe 658
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 313 CTGAATTGGCTGTGTCTCTCAATCCCTTTGGGCTTCGAATGTCTAGACCAACTTAT 372
Qy 659 GlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu 678
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 GGAGAACTCATATCTTTGGAGAGATAACAGTGAATCTACAGGGTGGGTATGTTTATTCTG 432
Qy 679 AsnArgPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAlaTyr 695
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 AATCCGCTACTTTTGGAAATCCTGATATAGCAGCTGAGTATAGACACCCACCTGTTCTCTC 492
Qy 696 ThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPhe 715
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 CTGTATAGAGATGCTCATGAAGAGCTTTGTCCAAAGTTTACATTTGAAAAGATTATTGTTG 552
Qy 716 LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeu 735
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 TTGCTCTGTTTCTTGTGATTATGCTTAAATTTCCAGACTCATGTATCATGATCCTTGTCTC 612
Qy 736 PheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGlu 755
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 TTCGTAAAGATGCGGAATTCAGCTAGTAAAGAAATCCTTTGGCTTTTTCAGGAT 672
Qy 756 LeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGln 775
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 673 TTCCTAAGTGGTGAAGGTGACCTTTCCGTCACCTTGGCTTTATTGGGATTACCTGTTAAC 732
Qy 776 HisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeuAlaValAspLeu 795
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 CATGTCAGACACCATTTGATGAATTTGCGGTACAAATCTTCGCGTAGACTTG 792
Qy 796 ArgAspGlyValArgLeuThrArgValValGluValLeuLeuLeuArgAspAspLeuThr 815
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 CAATGTGGAGTGGCCCTTGTCCGAACCATGAACTCTCTCACACAGAACTGGGACCTCTCA 852
Qy 816 ArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAla 835
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 853 AAGAAACTCAGGATTCGGCAATAAGTCGTCTTCAAAAGATGCACAAATGTTGACATGTT 912
Qy 836 LeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIleAla 851
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 912 CTTCAAGTCTTAAATCACGAGGAATTGAATTAAGTAGAGCATGGAAAATACAAATCTA 972
Qy 852 AlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIle 871
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 TCTAAGGATATTGGATAGCCACAGAAAAAACTCTCAGGTGCTTTGGAAAAATAGCG 1032
Qy 872 TyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArg 891
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1033 TTTGCTTTTCAG----- 1044
Qy 892 ArgHisTrpLeuHisValValIleGlnArgIleArgHisLysLysGluLeuMetArgArg 911
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1044 ----- 1044
Qy 912 HisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrVal 931
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1045 -----GTGATATTTCCTTAACTTAGATCAATTAAGAGAGAAAT 1086
Qy 932 LysLeuPheLys-ThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgAr 951
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1087 GCCTTTCTTAAACACACAAAAGA----- 1108
Qy 951 gTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTr 971
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1109 ---GTATAAGAAAAACAA---TATCTCTACTATCATCCATTTTGTGATC----- 1153
Qy 971 pTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAl 991
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1154 -----TTATTATAAGAAAAAGCAAGGAGATAGTGGTTCCTTTGACAAATAT 1203
Qy 991 aAla-----llePheLeuGlnArgIleTrpAr 1000
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



Db 1204 AGTGAACAAATATATTGATGCAATGGTAAATGCTGTTGGCTTCTTATAATAA 1263  
 Qy 1000 GARGArgLeuPheAlaLysLysLeuAlaAlaGluThrAlaArgLeuGlnArgse 1020  
 Db 1264 AAGTGGAGAAATTTACAGTGTCTT----- 1288  
 Qy 1020 rGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuG1 1040  
 Db 1289 -----TCTCAGACGGCGGTGTATGTTACCTGATCCACCAATTACC----- 1330  
 Qy 1040 yArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgAr 1060  
 Db 1331 -----ATCCCTGTGTATGTGCGCATTTGACGCTATATGTACGGTACT 1371  
 Qy 1060 gMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAl 1080  
 Db 1372 ACTCAA--CTGTGGAATGTA-----CGCAAACTGTTTCAGTGTATTAAATTC 1419  
 Qy 1080 aAlaIleAsnIleGlnGln-----ArgTrpArgAlaLysLeuSerMetArgLysCysAsnAl 1099  
 Db 1420 TCATCTGAATCTGATGACAGTCTCTCG----- 1447  
 Qy 1099 aAspTrpLeuAlaLeuArgSerSerValLeuLysVal-----GlnAlaTyrArgLysAlaTh 1118  
 Db 1448 -----ATATGCTCTCTTAAGCAATTTGATCATGAGAAATACTTCAGAGCTATACAAAGAGCTC 1503  
 Qy 1118 rIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLe 1138  
 Db 1504 CTAGAAATGAA-----AGAAAAATTTTCACTTGGTT 1536  
 Qy 1138 uGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLe 1158  
 Db 1537 AGGTCTGAGTTAGAGACCTTGTGGAA-----TACCTG----- 1570  
 Qy 1158 uArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAs 1178  
 Db 1571 -----CTATGATTAATCATTACAG----- 1588  
 Qy 1178 pArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGln---ArgArgTrpAr 1197  
 Db 1589 -----ATATGTCAAATACAAATTCAGATGAAAGGTGGTT 1623  
 Qy 1197 gAlaThr-----LeuGlnMetArgArgGluArgLysAs 1208  
 Db 1624 -ATTACCTATTTCATTCTTTGTGCAAGCTTTTGGATCTTCGTAAAGAAATAAGAGC 1682  
 Qy 1208 nTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLysArgG1 1228  
 Db 1683 TGCTCGACTCATCAAAACAATGAGAGAAATATAAACTMAAAACAGATCTCAAAACGCCA 1742  
 Qy 1228 uMetLysLysGlnArgAla-----GluPheLeu----- 1237  
 Db 1743 TCAGAGAGAGAGAAAGCTGCAGAAATATTCAATGGCTGTATCAATTTTCTAGCAA 1802  
 Qy 1238 -----GlnLeuLysLys-----ValThrLeuValValGlnLys----- 1248  
 Db 1803 ACAGAAGTTGAGAAAAGAGTTAATGCAGCACTGTCATTACAGAAATATTGGCGAAGAGT 1862  
 Qy 1249 -----ArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeuAr 1267  
 Db 1863 CTTAGCACAGAGAAATATTAAATGTTTAAAGAAAGAAAGCTGGAAGAAAGTTCAAAATAA 1922  
 Qy 1267 gGluValThrIle-----LysLeuGlnArgPheHisAlaGlnLysSerMetArgPh 1285  
 Db 1923 ACCAGATCCTATTTCAGGCTATGTGGAGAGATATAGACCAAGAAATATTATGTAA 1982  
 Qy 1285 eMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTrpAr 1305  
 Db 1983 AGTGAAGCTGCTGCAAG-----ATTCAAGCTGGTATAGATGT-----TGAG 2027  
 Qy 1305 gAsnHis-----LeuLe 1309

Db 2028 AGCACAAAGAAATATCTAGCTATATTAATAAGCTTTAAAAATTATTCAAGGTGCTTCTTA 2087  
 Qy 1309 uArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaIleThrLeuGlnAr 1329  
 Db 2088 TACCAAACTAGAGACAACACGCTTTTGAATGTGAGAGCATCAGCAATATTCATTACAGAG 2147  
 Qy 1329 gArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLys-- 1348  
 Db 2148 AAAATGGAGAGCTATCTCTCGAAAGATAGCTCATCAACACTTCTTTAATGATAAAG 2207  
 Qy 1349 ----GlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysG1 1367  
 Db 2208 ACATCGAGCTGCTTGTTCATCCACAGCACATTATAGAGATATAAGCA-----AGCA 2261  
 Qy 1367 nValValLeuTyrGlnLysGlnArgGluAlaIleIleLys-----ValGlnArgArgTy 1385  
 Db 2262 GGTCTTCTTCGCGAGAAATCTGCTGTGTATCATCACAAAATATATACGAGCCAGGA 2321  
 Qy 1385 rArgGlyAsnLeuMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaVa 1405  
 Db 2322 GGCTGGAAGCATGAAGATGATAATATATGAAATTT-----AAAAATCTAC 2369  
 Qy 1405 lIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuCysLysAlaGlyTy 1425  
 Db 2370 AGTTATCTTACAA----- 2382  
 Qy 1425 rArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAl 1445  
 Db 2383 -----GCATGGTGGTGGTGGCTAGTA----- 2406  
 Qy 1445 aArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaph 1465  
 Db 2407 -----CGAAAAAGATTTTGAACAGAGACCCAAATTCACCTTCTTCATCTCAC 2456  
 Qy 1465 eIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgArgAl 1485  
 Db 2457 TGCAGCTGCATATTATCACTCAAT----- 2481  
 Qy 1485 aAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAs 1505  
 Db 2482 -GCTGTAGAAATTCAAAGAGCCTATATAACTTTACCTGGCTGTGAAGAAATGCTACAGCA 2540  
 Qy 1505 pTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgse 1525  
 Db 2541 GGT-----AATTCACTCATCTGTTTACAGAGATGCTTTCGAGCAAGATTACA 2588  
 Qy 1525 rMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnG1 1545  
 Db 2589 AGAAAGAGATTTATTTCAGAAATATCATAGCATC----- 2622  
 Qy 1545 nLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuArgCy 1565  
 Db 2623 -----AAAAAGATTGAGCAGTCAAGATGCTCAGCCAGCGAAGATAG 2669  
 Qy 1565 sSerMetProGlyPheGlnAlaArgGlyPheMetAlaArgLysArgPheGlnAl 1585  
 Db 2670 GGCTGCATCAGTAATAACAGAAAGCAGTGCCTATTTCTCTCCGTAAAGAACAGCAAAA 2729  
 Qy 1585 aLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleG1 1605  
 Db 2730 ATTCACTAGT-----GGATCATTA-----ATTCA 2756  
 Qy 1605 nArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeuAspI1 1625  
 Db 2757 GGCAATTATGGAGAGCTATTCTTGGAGGAAGAAATGATTGTACAAAAATTAAGCTAT 2816  
 Qy 1625 eArgLysArgIleAlaGlnLeuArgGlnGlnAlaLysAlaValAsnSerValArgCysLy 1645  
 Db 2817 ACAGCTAAGTCTTCAGTTGTTAATAGGAGATTCGAGAGAGAAACAACTCTACAAAAG 2876  
 Qy 1645 sValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaVa 1665  
 Db 2877 AACTGCACCTTGCACTTCATTACTTTTGACATATAAGCACCTTTCTGCCATTTCTGAGGC 2936





754	QY	SerGluLeuLeuAlaAenIleGlyAspIleThrArgGluLeuArgLeuGlyTyrVal	773
723	Db	CGAGATTCCTAAGTGTGAAGGTGAACCTTCGCCGTGCCTTGGCTATTATGGGATACCT	782
774	QY	LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal	793
783	Db	GTAAACCATGTTCCAGACACCATTTGATGAATTTGATTTCCGGTTACAAATCTTCGGCTA	842
794	QY	AspLeuArgAspGlyValArgLeuThrArgValValGluValLeuLeuLeuArgAspAsp	813
843	Db	GACTTGCAATGTGGAGTGCCTTGTCCGAACCATGGAACCTTCTCACAGAACTGGGAC	902
814	QY	LeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLys	833
903	Db	CTCTCAAGAAACTCAGGATTCGGCAATAAAGTCGTCTTCAAAAGATGCACAATGTTGAC	962
834	QY	LeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu	849
963	Db	ATTGTTCTTCAAGTTCCTAAATCAGAGGAATTGAATTAAGTGAAGCATGGAATACACA	1022
850	QY	IleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTpgLys	869
1023	Db	ATTCTATCTAAGGATATTGGATAGCCACAGAGAAAACCTCTCAGTTGCTTTGGAAA	1082
870	QY	LeuIleTyrLysPheArg	875
1083	Db	ATAGCGTTTCTTTTCAGTGGATATTTCCTTAACTTAGATCAATTAAGGAAGAAATT	1142
875	QY		875
1143	Db	GCCTTCTAACAACACAAAGATATAAGAAACAATATCTCTACATCATGCCATTCT	1202
876	QY		880
1203	Db	GATGATCTTATTAAAGAAAAAGGCAAGGATAGTGGTTCCTTTGACCAATATAGT	1262
881	QY		888
1263	Db	GAATAACATAAAGTTATTGATGGATTGGGTAAATGCTGTTTGGCTTCTTATAATAAAG	1322
888	QY	sT-tp	889
1323	Db	GTGGAGAAATTTACAGTGCTCTTCTCAGCGCGGTGTGTATGTTACCTGATCCACAT	1382
889	QY		889
1383	Db	TACCATCTTGCTATGTGCCATTAGACGCTATATGTGTCAGCGTACTACTCAAACTGGAA	1442
890	QY		890
1443	Db	TGTACGCAAACTGGTTCAGTGGTATTAAATTCATCATCTGAATCTGATGACAGTTCCTG	1502
890	QY	pArgArgHisTrpLeuHisValVal	898
1503	Db	GATATGCTACTTAAGCATTTGATCATGATAAATPACTTCAGAGCTATACAAAGAGCTCCTA	1562
899	QY		905
1563	Db	GAATAAGAAAGAAATTTT-CACITGGTGTAGTCTGCAGCTTAGAGACCTTGTTGGAAT	1621
905	QY		905
1622	Db	ACCTGCTATGATTAAATCATTCAGATATGTCAAATAACAATCCAGATGAAAGGTGGTTAT	1681
906	QY		915
1682	Db	TACCTATTTGTCATTTCTTTGTGCAAGGCTTTTGGATCTTCGTAAGAATAAGAGCTGC	1741
915	QY	aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLys	935
1742	Db	TCCACTCATACAACACATGGAGAAATATAAATAAACAAGATCTCAACGCCCATCA	1801
935	QY	sThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyrLeuAlaGln	955

Db	1802	G	----	GAGACAGACAGAAAGCTGCACAAATATTATTCATTTGGCTGTAATCAATTTTCTAGCAAA	1858
Qy	955	n	YsglnLeu	TyrGlnSerTyrHisSerIleIleGlnArgTrpTrpArgAlaG1	975
Db	1859	ACAAGATTGAGAAAAGAGTTAATCAGCACCTCGTCATCTCAGAAAATATTGG	1910		
Qy	975	n	GlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLe	995	
Db	1910	-----	-----	-----	1910
Qy	995	u	GlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl	1015	
Db	1911	-----	-----	-----	1911
Qy	1015	a	ArgLeuGlnArgSerGlnLysGlnAlaAlaAlaSerTyrIleGlnMetGlnTrpAr	1035	
Db	1955	AAAGCTGGAAAGATTCAAAATAAA	2008		
Qy	1035	g	ThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe	1055	
Db	2009	AAAGATATCC	2056		
Qy	1055	t	PheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1	1075	
Db	2057	CATCCTCGAATCTAGGATAAGA	2104		
Qy	1075	n	GlnLeuLysArgAlaAlaIleAlaIleGlnArgTrpArgAlaLysLeuSerMetAr	1095	
Db	2105	ACGATATCTTTGGGTACAGTACAATTCAGAGCATTTGGCGTCTTATTAAAGAGAAA	2164		
Qy	1095	g	LysCysAsnAlaAspTyrLeuAlaIleArgSerValLeuLysValGlnAla	1114	
Db	2165	ACAAGATCAACAAGATATGAATGCTAAATCATCACTCTTATATCAATCTATGTT	2224		
Qy	1114	r	ArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs	1134	
Db	2225	CAGAAAA	2246		
Qy	1134	n	ValIleCysLeuGlnGlnArgLeuArgAla	1153	
Db	2247	-----	-----	-----	2296
Qy	1153	u	AsnTyrLeuArgLeuArg	1169	
Db	2297	ATGCAATTTAAGAAACAAGCTAAGAAAGAAATCTGCTATTATCATCAATCATGGTA	2356		
Qy	1169	r	ArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysI1	1189	
Db	2357	TAGAATGCATAAAGAATTA	2404		
Qy	1189	e	lleAsnValGlnArgArgTrpArg	1197	
Db	2405	TGTTATCATTCAGAAAAGATTTCGG	2429		
RESULT	10				
BD156229					
LOCUS	BD156229		2453 bp	DNA	linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD156229				
VERSION	BD156229.1				
KEYWORDS	JP 2002191363-A/11072.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Fukuyoka; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 2453)				
TITLE	Oca, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakanatsu, A., Nagai, K. and Otsuki, T.				
JOURNAL	Primer for synthesizing full-length cDNA and use thereof				
COMMENT	Patent: JP 2002191363-A 11072 09-JUL-2002; HELIX RESEARCH INSTITUTE OS Homo sapiens (human)				



Qy 915 aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheIy 935  
 Db 1742 TCGACTCATCAACAACATGAGAGAAATATATAAATAAAGACAGATCTCAAAAGCCATCA 1801  
 Qy 935 sThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgAspGlyLeuAlaG1 955  
 Db 1802 G---GAGAGAGAGAAAGTTCAGAAATTTCAATGGCTGTAATCAATTTCTAGCAAA 1858  
 Qy 955 nLysGlnLeuTyrGlnSerTyrHisSerIleThrLeuGlnArgTyrTrpArgAlaG1 975  
 Db 1859 ACAGAAGATTGAGAAAAGAGTTAATGAGCAGCTCTGTCATTGAGAAATATGG----- 1910  
 Qy 975 nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLe 995  
 Db 1910 ----- 1910  
 Qy 995 uGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015  
 Db 1911 -----CGAAGAGTCTTAGCAGAGAAATTTAATGTTAAAGGA 1954  
 Qy 1015 aArgLeuGlnArgSerGlnLysGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035  
 Db 1955 AAAGCTGAGAAAAGTTCAAAATAA-----GCAGCATCACTTATTCAGGGATATTCGAG 2008  
 Qy 1035 gThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055  
 Db 2009 AAGATATTCC-----ACTAGACAAAGATTTCTGAAATTTGAAATATTATTCAAT 2056  
 Qy 1055 tPheValGlnArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1 1075  
 Db 2057 CATCTGCAATCTAGATAAGA-----ATGATAATTGCTGTACATCTTATAA 2104  
 Qy 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysSerMetAr 1095  
 Db 2105 ACGATATCTTTGGGCTACAGTTACAAATTCAGAGGCAATGGCGTCTATTATAAGAGAAA 2164  
 Qy 1095 GlyCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAla---Ty 1114  
 Db 2165 ACAGATCAACAAGATATGAAATGCTAAATCTTAAATCTTAAATCTTAAATCTATGTT 2224  
 Qy 1114 tArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
 Db 2225 CAGAAAA-----TGGAGCAACGTAAA--- 2246  
 Qy 1134 nValIleCysLeuGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgG1 1153  
 Db 2247 -----ATGCAATCACAAGTAAAGCTACAGTAATATTGCAAGAGCTTTTAGAGA 2296  
 Qy 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169  
 Db 2297 ATGGCATTTAAGAAAACAGATTAAGAGAAATTTCTGCTATTATATACATCATGTTA 2356  
 Qy 1169 tArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysI1 1189  
 Db 2357 TAGATGATCAAGAATTA-----CGAAGTATATTATATTAGATCTTGCT 2404  
 Qy 1189 eIleAsnValGlnArgArgTrpArg 1197  
 Db 2405 TGTATCATTCAGAAAAGATTTTCGG 2429  
 RESULT 11  
 LOCUS AK001380 2453 bp mRNA linear PRI 01-AUG-2002  
 DEFINITION Homo sapiens cDNA FLJ10518 fis, clone NT2RP2000814.  
 ACCESSION AK001380  
 VERSION AK001380.1 GI:7022604  
 KEYWORDS oligo capping; (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1

AUTHORS  
 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
 Nishikawa, T., Negai, K., Sugano, S., Shiratori, A., Sudo, H.,  
 Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,  
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,  
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,  
 Nakamura, Y., Nagahara, K., Masubo, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 JOURNAL  
 REFERENCE  
 2 (bases 1 to 2453)  
 Isogai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@helix.co.jp, tel: 81-438-52-3975, fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection;  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 Location/Qualifiers  
 1. 2453  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP2000814"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP2"  
 /note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal  
 precursor cells after 2-weeks retinoic acid (RA)  
 induction."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9,398-38 Length: 2453  
 Score: 834.00 Matches: 235  
 Percent Similarity: 44.65% Conservative: 153  
 Best Local Similarity: 27.04% Mismatches: 228  
 Query Match: 8.77% Indels: 254  
 DB: 9 Gaps: 22  
 US-09-914-698-1 (1-1861) x AK001380 (1-2453)  
 Qy 522 AspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTyrLeuAsnAlaVal 541  
 Db 6 GATGAACGCTGGAAGGAAAGAGCAGGACAGGGCTTCACTTGGTGGTTAAATTTATATTA 65  
 Qy 542 SerIleProAlaAspLeuAspAlaAspLeuAan---AsnLysIleAspValGlyLysLeu 560  
 Db 66 ACC---CCTGATGACTTCACTGTAAAAACAATAATTTCTGAAGTAAATGCTGCTACTCT 122  
 Qy 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577  
 Db 123 CTTTGGGAATAGAGAAATCAACATATAAATAGTGTTCCTAGAGCACCCTACAAAGAGGA 182  
 Qy 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaVal 596  
 Db 183 ATGCTCTCAGAGCTTATACCTCGGTGTAGTTTAAACAGACTACGTCGTGAGCATGC 242  
 Qy 597 GluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616  
 Db 243 CGTTTGGTTTCTCTGAAAAAATGTTAAAGCTATTAAAAAGCTTGAATTTGAAATTTGAA 302  
 Qy 617 LysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArg 636  
 Db 303 GCTAGCGGTAAATTTGTCGAAAAATAGACACCATTCGAAAGATGTGGGAGAGAGCTCAG 362  
 Qy 637 ThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuVal 656  
 Db 363 AAAGTCCTGAATTTGGCTGTTCCTACATCTTTTGTGGCTTCGAATTTGGTTCAGAGACA 422

QY 657 ValPheGlyGluIysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676  
 Db 423 ACTATGCGAAGACTCATCTTTGGAGAGATACAGTGTATGTCACAGGTTGGCTATGTTT 482  
 QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693  
 Db 483 ATTCTGAATCGCCTACTTTTGGAAATCCGTATATAGCAGCTGAGTATAGACACCCACAGTTT 542  
 QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysIysHisSerLeuGlnLysIle 713  
 Db 543 CCTCACCTGTATAGATGTGTCATGAAGAAGCTTTGTCCAAAGTTTACATTTCAAAAGTTA 602  
 QY 714 LeuPheLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733  
 Db 603 TTGTTGTGTCGTCTTCTTGATATGCTAAATTTCCAGACATCATGATCAATGATCCT 662  
 QY 734 CysLeuPheValLysIysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 753  
 Db 663 TGTCTCTTCTGTAAGATGCGGAATTCAGGCTAGTAAAGAAATCCTTTAGGCTTTTCA 722  
 QY 754 SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrVal 773  
 Db 723 CGAGATTCTTAAGTGTGAAGTGCACCTTTCCCGTCGCTTGGCTTATTGGGATACCT 782  
 QY 774 LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal 793  
 Db 783 GTTAACCATGTTGAGACACCATTTGATGATTTGATTTTGGCGTTACAAATCTTGGCGTA 842  
 QY 794 AspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAsp 813  
 Db 843 GACTTGCATGTGAGTGCCTTGTCCGAACCATGGAATCTTCTCACACAGAACTGGGAC 902  
 QY 814 LeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLys 833  
 Db 903 CTCTCAAGAAGACTCAGATTCCGGCAATGAGTGTCTTCAAGAGATGSCAATGTTGAC 962  
 QY 834 LeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAsp 849  
 Db 963 ATTGTTCTTCAAGTCTTAAATCACAGGAGATTGAATTAAGTATGATGATGAGAAATACA 1022  
 QY 850 IleAlaAlaGlnAspIleValAspGlyHisArgGluIysThrLeuSerLeuLeuThrGln 869  
 Db 1023 ATTCTATCTAAGGATATTGATGATGACGACAGAGAAAAAACTCTCAGGTTGCTTTGGAAA 1082  
 QY 870 LeuIleTyrLysPheArg----- 875  
 Db 1083 ATAGCGTTTGTCTTTCAGTGGATATTTCCCTTAACCTTAGATCAATTAAGAGAGAAATTT 1142  
 QY 875 ----- 875  
 Db 1143 GCCTTTCTAAACACACAAAGAGTATAAGAAAAACATATCTCTACTATCATGCAATCT 1202  
 QY 876 -----SerProLysPheHis----- 880  
 Db 1203 GATGATCTTATTAAAGAAAAAGGCAAAAGGATAGTGTCTTCTTGAACAATATAGT 1262  
 QY 881 -----AlaAlaAla-ThrValLeuGlnLys 888  
 Db 1263 GAAAAACAATAAGTTATTGATGGATTGGGTAAATGCTGTTTGTGCTTCTTATAATAAAG 1322  
 QY 888 sTrp----- 889  
 Db 1323 GTGAGAAATTTACAGTGTCTTTCTCAGACGGCGGTGTATGTATTACCTGATCCACCAT 1382  
 QY 889 ----- 889  
 Db 1383 TACCATCTTGTCTATGTGCCATTAGACGCTATATGTGACGGTACTACTCAAACTGTGGAA 1442  
 QY 890 -----Tr 890  
 Db 1443 TGTACGCAAACTGGTTCAGTGGTATTAAATTCATCTCAATCTGATGACAGTTCTCTG 1502

QY 890 pArgArgHisTrpLeuHisValVal----- 898  
 Db 1503 GATATGTCATCTTAAAGCAATTTGATCATGAAATACTTTCAGAGCTATACAAAGAGCTCCTA 1562  
 QY 899 -----IleGlnArgArgIleArgHis----- 905  
 Db 1563 GAAAAATGAAAGAAAAATTTT-CACTTGGTTAGGTCGTCAGCTTAGAGACCTTGGTGAAT 1621  
 QY 905 ----- 905  
 Db 1622 ACCTGCTATGATTAATCATTCAGATATGTCAAATACAAATCCAGATGAAAGGTGGTTAT 1681  
 QY 906 -----LysGluLeuMetArgArgHisArgAlaAl----- 915  
 Db 1682 TACCTATTTTGCATCTTCTTTGTGCAAGGCTTTTGGATCTTCGTAAAGAAATAGAGCTGC 1741  
 QY 915 aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLys 935  
 Db 1742 TCGACTCATACAAACACATGAGAGAAATATAAACTAAAAACACAGATCTCAAAOCGCCATCA 1801  
 QY 935 sThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyrLeuAlaG1 955  
 Db 1802 G--GAGAGAGAGAGAGCTGCAAGATTAATTCATTTGGCTGTGTAATCAATTTCTAGCAAA 1858  
 QY 955 nLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaG1 975  
 Db 1859 ACAAGATTGAGAAAAAGAGCTTAATGCAGCACTCGTCAATTCAGAAATATTGG----- 1910  
 QY 975 nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaIlePheLe 995  
 Db 1910 ----- 1910  
 QY 995 uGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaAlaGluThrAl 1015  
 Db 1911 -----CGAAGAGTCTTAGCAGAGAGAAATTAATATGTTAAAGAAAGGA 1954  
 QY 1015 aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035  
 Db 1955 AAAGCTGGAATAAGTTCAAAATAAA-----GCAGCATCACTTATTTCAGGGATATTGGAG 2008  
 QY 1035 gThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055  
 Db 2009 AAGATATATCC-----ACTAGACAAAGATTTCTGAAATTTGAAATATTATTCAAT 2056  
 QY 1055 tPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1 1075  
 Db 2057 CATCTGCAATCTAGGATAAGA-----ATGATAATTGCTGTTTACATCTTATAA 2104  
 QY 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetAr 1095  
 Db 2105 ACGATATCTTTGGGCTACAGTTACAAATTCAGAGGCAATGGCGTCTTATTTAAGAAGAA 2164  
 QY 1095 gLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAla---Ty 1114  
 Db 2165 ACAAGATCAACAAGATATGAAATGCTTAAATCATCAACTCTTATATCAATCTATGTT 2224  
 QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
 Db 2225 CAGAAAA-----TGGAGCAACAGTAAA-- 2246  
 QY 1134 nValIleCysLeuGlnGlnArgLeuArgAla-----IleMetLysMetArgGluGlnArgG1 1153  
 Db 2247 -----ATGCAATCAACAAGTAAAAAGCTACAGTAATATTGCAAGAGCTTTTAGAGA 2296  
 QY 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169  
 Db 2297 ATGCGCATTTAAGAAAAACAGCTAAAGAGAAATTTCTGCTATTATCATACAAATCATCGTA 2356  
 QY 1169 rArgMetArgGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysI1 1189  
 Db 2357 TAGAATGCATAAAGAAATTA-----CGGAAGTATATTTATATTAGATCTTCTGT 2404  
 QY 1189 eIleAsnValGlnArgArgTrpArg 1197



Db	1674	TCCTGCAGTTTCAGATATTTCTTAGTCATAGCCACAATAAACCTTAAGAGACGTCCAATA	1733
Qy	394	-----HisLysArgArgSer-----	398
Db	1734	CTTTCTGCCACTGTTACTTAAAGGAAGGCCACCTGTACAGAGAAAACCAAACTGAGATT	1793
Qy	399	-----HisGluLeuSerPhe	403
Db	1794	AATAAACCAAAAGCAAAAGATCTCAACAGTGCAGTGGGTGAACATAAGAAAGTAATA	1853
Qy	404	SerAspAlaProSerAsnGluSerLeu-----	412
Db	1854	AATATCAAAAGAAAGAAAGAGATTTTCATTCTTATCTTCCAAATTATAGATCCAATATTA	1913
Qy	413	-----TyrArgAsnGlu-----	421
Db	1914	AGTAAATCTAAGAGTTATATAAAACGAGGTAAACACCTCTTCGCAACACAGCTTCAGTGT	1973
Qy	422	ProProLysLysGln-----ArgValGluAspThrThrLeuProArgSerAlaAlaPro	439
Db	1974	CGGAAAGAAAGACGCGATGGAAGATGCGAAATGTG---AGAGTTGCAATTACA	2030
Qy	440	AlaAsnAlaSerAlaArgSer---SerSerAlaHisAlaTPrpProHisAlaGln-----	456
Db	2031	GAACATACAGAGTGCAGAGAATCAAAAGAAATCCATTTCCTCCCTCAGAGCCTAAACA	2090
Qy	457	-----SerLysLysPheLysLeuAla	463
Db	2091	TCAGCTGTTAAGAAAAACAAAAATGTGACAACACCCATCTCTCAAAACGTTATTAGCAACAGA	2150
Qy	464	GlnThrMetSerLeuMetLysLysPro-----	478
Db	2151	GAGAAATTAACCTGGAAGAGAAACCTGAATTTATCAATATTCAGAACTCCAAATTTCTAAA	2210
Qy	479	ArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyr-----	495
Db	2211	ACAACAAAGGACAAACCAATTCATCGCTGTGCACAGTCCAGTTTTCCTTCATAAAA	2270
Qy	496	---MetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThrIleAspProPhe	514
Db	2271	CCATTAAAAACAGATATTCACAGACACCCGATG-----CCATTT	2309
Qy	515	LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLys	534
Db	2310	GCTCCAAAAACATGTTTATGATGACGCTGGGAAGGAAAGCGAGGACAGGCGTTCACT	2369
Qy	535	LysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsn---Asn	553
Db	2370	TGGTGGTTAAATTTTATTTAAACC---CCTGATGACTTCATGCTGATAAAACAAATATTCT	2426
Qy	554	LysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValVal-----	571
Db	2427	GAAGTAAATGCTGCTACTCTTCTTTTGGGAATAGAGAATCAACATAAAATAGGTTCCT	2486
Qy	572	---AlaProThrLysGluGluGlnProSerMetAsn---TyrLeuThrLysTyrArgLeuGlu	589
Db	2487	AGAGACCTCTCAAAAGAGGAATGTCTCTCAGAGCTTATATCTGCTCGGTAGGTTAAAC	2546
Qy	590	ThrLeuArgLysAlaAlaValIleGluLeuPhePheSerGluGlnMetArgLeuProCysSer	609
Db	2547	AGACTACGTCGTGCAGCATCGCTTGTCTGTTACTCTGAAAAAATGGTTTAAAGCTATTAAA	2606
Qy	610	LysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHis	629
Db	2607	AAGCTTGAATTTGAATTTGAAGCTAGGCGGTAAATGTTTCGAAAGATAGACACTATGG	2666
Qy	630	LeuAspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTrp	649
Db	2667	AAAGATGTGGGAGAACGTCAGAAAGTCTGATTTGGCTGTTGTCTCTACATCCCTTTGGG	2736
Qy	650	LeuArgLeuGlyLeuGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAsp	669

Db	2727	CTTCGAATTGGCTAGAGACCACTTATGGAGAACTCATATCTTTGGAGAGATAACAGTGAT	2786
Qy	670	IleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArgAsn-----LysCys	686
Db	2787	GTACAGGGTTGCTATGTTTATCTGAATCGCCTACTTTTGGAAATCCTGATATAGCAGCT	2846
Qy	687	GluGlnGlnArgTyrSerIysAlaTyrThrLeuThrGluGluTyrAlaGluThrIleLys	706
Db	2847	GAGTATAGACACCCCACTGTCTCCTCTATAGAGATGGTCAATGAAGAGCTTTGTCC	2906
Qy	707	LysHisSerLeuGlnIysIleLeuPheLeuPheProPheLeuAspGlnAlaLysGlnLys	726
Db	2907	AAGTTTACATTTGAAAGATTTATTGTTGGTCTGTTTCTTGAATATGCTATAAATTTCC	2966
Qy	727	ArgIleValIysHisAsnProCysLeuPheValIysLysSerProHisLysGluThrLys	746
Db	2967	AGACTCAATTCATGATCCCTTGTCTTCTGTGAAGATCCGGAATCAAGGCTAGTAAA	3026
Qy	747	AspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGlu	766
Db	3027	GAATTCCTTTTGGCTTTTTCACGAGATTTCTTAAGTGGTGAAGTGACCTTTCCCGTAC	3086
Qy	767	LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr	786
Db	3087	CTTGGCTTTATGGGATTAACCTGTTAAACATGTTCCAGACACCAATTGATGAATTTT	3146
Qy	787	AlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGlu	806
Db	3147	GCGTTTACAAATCTTGGCGTAGACTTGCAATGTGGAGTGGCCCTTGTGCGAACCACTGAA	3206
Qy	807	ValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeu	826
Db	3207	CTTCTCACAGAACTCGGACCTCTCAAGAGAACTCAGGATTCGGCAATTAAGTCGTCCT	3266
Qy	827	GlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu	846
Db	3267	CAAAAGATGCACAAATGTGACATGTTCTTCAAGTTCTTAATACACGAGGAATTGAATTA	3326
Qy	847	-----GlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGlnLys	862
Db	3327	AGTGATGAGCATGGAATACAACTTCTATCAAGGATATTGTGATAGGCACAGAGAAAA	3386
Qy	863	ThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArg	875
Db	3387	ACTCTCAGTTGCTTTGGAAAAATAGCGTTTCTGCTTTTCAG	3425
RESULT 13			
AC097078/c			
LOCUS			
DEFINITION			
AC097078			
AC097078.7			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			



Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gregoire,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kwis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorenshewa,L., Louissegh,H., Lozano,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mapua,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwackelmech,O., Okwuonu,G., Olarinmoye,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,D., Primus,E., Pu,L.-L.,  
 Plopper,F., Poindexter,A., Popovic,D., Reeves,K., Regier,M.A., Reigh,R.,  
 Puazo,M., Quiroz,J., Rachin,E., Reuter,M., Richards,S., Riggs,F.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajda,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,E., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,X., Zhou,J., Zhou,S., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstein,G., and Gibbs,R.A.

# TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Unpublished  
 2 (bases 1 to 235179)  
 Worley,K.C.

Direct Submission  
 Submitted (09-OCT-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235179)  
 Rat Genome Sequencing Consortium.

Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24956330.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly ('contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGMW

Center clone name: CH230-83E5

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 228193 bases at least Q40  
 Consensus quality: 230105 bases at least Q30  
 Consensus quality: 231450 bases at least Q20  
 Estimated insert size: 238014; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 4 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

\* 1 37618: contig of 37618 bp in length  
 \* 37619 37718: gap of unknown length  
 \* 37719 228862: contig of 191144 bp in length  
 \* 228863 228962: gap of unknown length  
 \* 228963 232645: contig of 3683 bp in length  
 \* 232646 232746: gap of unknown length  
 \* 232746 235179: contig of 2434 bp in length.

FEATURES  
 Location/Qualifiers  
 1..235179  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-83E5"  
 1..1825  
 /note="wgs\_contig"

misc\_feature  
 source

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,52e-30 Length: 235179  
 Score: 753.00 Matches: 274  
 Percent Similarity: 43.14% Conservative: 198  
 Best Local Similarity: 25.05% Mismatches: 324  
 Query Match: 7.91% Indels: 298  
 DB: 2 Gaps: 39

US-09-914-698-1 (1-1861) x AC0907078 (1-235179)  
 Qy 873 LysPheArgSerProLysPheHisAlaAlaThrValLeuGln-----LysTrp 889  
 Db 159339 AGATTGAGA-----AAGCTGAATCTGCATCCCTGTGTAATCCATCTGTGTCAGACGG 159286  
 Qy 890 TrpArgArgHisTrpLeuHis-----ValValIleGlnArgArgIle 903  
 Db 159285 TGGAGCGACGCAAGCTGCAGCTCCAAACCAAGCGCGGTGCATTCGACGACGCTTTT 159226  
 Qy 904 ArgHisLysGluLeuMetArg-----ArgHisArgAlaAlaThrValIleGlnAlaVal 921  
 Db 159225 AGAAGATGCGGTGTCAGGAAACAGACACAGAGAAAGCGCGCTGTGTGATACAGTCGTGG 159166  
 Qy 922 PheArgGlyHis---GlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
 Db 159165 TACAGAATCCACAGGGAATTGCGAAGTACATTACATT-----AGGTCT 159121  
 Qy 941 AlaAlaIleIleGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960  
 Db 159120 TGTGTTGTAGTCATCCAGAGAAGAGTTCGTGTCCTTTCAGCCCAAAAGTATGTAAAGAGA 159061  
 Qy 961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
 Db 159060 AGGAAAGACGCTGCTCTGACTCTCCAGAGACACTTCAGGGCCCGTCAGATGGGAAGTTA 159001  
 Qy 981 HisArgGlnArgPheValGlnLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArg 1000  
 Db 159000 GCGCGACTGACTATTTTTCGAGAGCGCGCTGCATTCACACTGCAG----- 158953  
 Qy 1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
 Db 158953

Db 158952 -----GCTGCGCTTCAGGGCGCATGAAGGCTCGCCGTTCA 158920  
Qy 1021 GlnLysGlnGln---AlaAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeu 1039  
Db 158919 TACGGTTTGCAATTCAGAGCTGCTGTGTCTTCAGTCTCTATTTGGAGAATGAGACAG--- 158863  
Qy 1040 GlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArg 1059  
Db 158862 GGAAGAGTT-----AAATTTCTTAAATCTTTAAAGAGAGCTGTATCAAAATTCGAAGCA 158812  
Qy 1060 ArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArg 1079  
Db 158811 CACATAAGAAAAACAC-----CAACAACACTACAGAAATATCAGAGATAAGAAA 158764  
Qy 1080 AlaAlaAlaSerIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAla 1099  
Db 158763 GCAGCTGTTCACAAATTCAGACTCATTTCCGAGCTTCCATTTCCAGCCAGGAAGAGTCTAGCA 158704  
Qy 1100 AspTyrIleAlaLeuArgSerSerValLeuLysValGln---AlaTyrArg----- 1115  
Db 158703 TCTTACAGAAACACAGCTCTTCAGTCAATTTGTTCTACAGTCTGCATATACAGGATGCA 158644  
Qy 1116 -----LysAlaThrIleGln 1120  
Db 158643 GCCAGGAAAAACATTCGTCGTGCTTTAAACATCTGTAGTAAAGATTTCAGTCACTGTTATCG 158584  
Qy 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
Db 158583 GCTTTATCTCCAGAAAGAGTTTCGAGCTTGAGGAATGCTGCAGATA----- 158536  
Qy 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
Db 158535 AAGCTGCAGTCAATTCGTAATGAAACAGTACAGTAAAGCAGTATTTACAGTAAAGGCC 158476  
Qy 1161 AlaSerIleLeuValGlnLysArgTyrArg-----MetArgGlnGlnMetIleGlnAsp 1178  
Db 158475 GCTGCGGTATTTTACAGCGGTGTTACCGTTCCTCCAAAAGAGAAAGAGTCTGACAGGTG 158416  
Qy 1179 ArgAsnAlaTyrLeuArgThr-----ArgLysCysIleIleAsnValGlnArg 1194  
Db 158415 CGGAAAGCTGTATCAAACTGCAGTCTGCTTACAGAGATGCTGCTCAGAAAGCAGATG 158356  
Qy 1195 ArgTrpArgAlaThr-----LeuGlnMetArgArgGluArg 1206  
Db 158355 AGCTCCAAAGCAGAGCTGCCATTTCTCTGCAGTCTTACTTCAGGATGTGAGAGTGAGG 158296  
Qy 1207 LysAsnTyrIleHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLys 1226  
Db 158295 CAGCGTACTCTGAAGCGCGCAGCGGCTCTTGTCTATTCACACTTCTACAGTGCCTAC 158236  
Qy 1227 ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVal 1246  
Db 158235 CGAGCACAAGTCCATCAGAGGAAGATCTTCTGCGAGTCAAGAGGACAGCTGTTGTTTG 158176  
Qy 1247 Gln-----LysArgArgArgAlaLeuLeuGlnMetArgLys----- 1258  
Db 158175 CAAGCAGCTTACAGAGCGCGCAGAGCAGCGCGGCAAAATCAAGCAGAGTCTACAGCTGCC 158116  
Qy 1259 -----GluArgGlnGluTyrLeuHisLeuArg 1267  
Db 158115 GTTACAATTCAGAGGTTTTTAGAGGCCACAGTCTCAGAGATGCAATACAGACTGTGCTT 158056  
Qy 1268 GluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArg 1287  
Db 158055 CAGCTGTGTGAGGTTCAGAGATGGTACAGGGGACACAAAGTTGCTTCTGACATGAGA 157996  
Qy 1288 AlaLysTyrArgGlyThrGlnAlaAlaVal-----SerCys----- 1299  
Db 157995 ATTCACCTTTCTAAAGACAAGGGAGGCTGTGCTGTTCTGCACTCTGCTGCTGCTGCTG 157936  
Qy 1299 ----- 1299  
Db 157935 CAAGTTGGCAGCAGCTCAGGAGGCAACATCAGGCTGCAGTGAAGATCCAGGCTGCTTTC 157876

Qy 1300 -----LeuGln 1301  
Db 157875 AGAATGGCCAGGGCCAGAGGCGAGTATAAACTAGTGAGAAATGCTGCAGTCTCATCCAG 157816  
Qy 1302 MethIstTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArg 1321  
Db 157815 CAGCATGTTCAGGGCAGCGGGCTGCAGAGAGAGGCGACACTCTCGGAGTACATCCGCTCCGT 157756  
Qy 1322 GlnAlaIleThrLeuGlnArgArgTyrArgAlaArgLeu----- 1335  
Db 157755 CACGCCCGCTGTGTGTTTCAGGCTACCTGGAAGGGAAGATGCTGAGGAGAGATCGCA 157696  
Qy 1336 -----AsnMetIleLysGlnLeuLys 1342  
Db 157695 CAGCGGCACACAGTCCGGCGGCTCATACAGTCGCTACTACAGAATGTACAGTCCAGCTAG 157636  
Qy 1343 SerTyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLys 1362  
Db 157635 AAGTGGAGCTTCATGAAGAACCGCTGCTCGTCAGATTTCAGTGTGTATAGGGCTACAGA 157576  
Qy 1363 AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleLysValGln 1382  
Db 157575 GTTGGAAAGGAACAGCGCCATTTGTACTTGAACAAAGGCGCTGTAGTAAATTTTACAG 157516  
Qy 1383 ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIle----- 1395  
Db 157515 TCAGCTTACCCTGTGT-----ATGAAGCGAGGAAGACAGATAACTGAATGCCACAAAGCAGCA 157459  
Qy 1396 -----GluValTyrGlnLysGln-----Arg 1402  
Db 157458 GTCACTATACAGTCTAAATTTAGGGCTACAGAACACAGAAAGAAATATACACCTGTAGA 157399  
Qy 1403 GlnAlaValIleArgLeuGlnLysTrpArgSerIleArgAspMetArgLeuCysLys 1422  
Db 157398 ACTTCTCTCTATTAATTCAGAGATGGTATCGGAATATTAAATATACCACCCAGCAACGT 157339  
Qy 1423 AlaGlyTyrArgArgIleArgLeuSerLeuSerIleGlnArgLysTrpArgAlaThr 1442  
Db 157338 GAGCATACCTTACCTTAAAGAGAGCAGCAACTAAAGTCCAACTGTGTACAGAGGC--- 157282  
Qy 1443 ValGlnAlaArgArg----- 1447  
Db 157281 ATTCAGTAAGGAGACACATTCAGCACATGTACATGCGCGCCACACTCATTTAAAGCCTTG 157222  
Qy 1447 ----- 1447  
Db 157221 TTTAAGATACAGCAGTCGACAGCAGCGGCTGTGATCCAGGTGAGGTACAGAGCTATTAT 157162  
Qy 1448 -----GlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
Db 157161 CTAGGTAAATTCAGCAGGAAAGATGACTGACGACTTTTAGGGCTATTAAAGACCTTCAA 157102  
Qy 1464 AlaPheIleArg-----AlaThr 1469  
Db 157101 GCCGTGTACAGAGCAGCAAGAGTTAGAACACTTTAAGAAAGATGATTTCCGACGCAAG 157042  
Qy 1470 LeuLeuMet-----ArgGlnGlnArgArgGluPheGluMetLysArgArg 1484  
Db 157041 CTCATTTCAGTCATCTTTAGGAGACACAGACAGCAACATATTTCCACCGCTTGAGGAAG 156982  
Qy 1485 AlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGln 1504  
Db 156981 GCAGCCACAGTGTGACGAGAGTACCGAGCAGTAGTAAGAGGAAGAGTCTCAGTTTCAG 156922  
Qy 1505 AspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPheArg----- 1521  
Db 156921 AGGTACAGCAGACTGAGGCGTCTGTAACTCTGTCAGGCTGCTTTAGGGGACTGAA 156862  
Qy 1522 AlaAsnArgSerMetLysGln----- 1528  
Db 156861 GCGCGGAGACACTTAAAGACTATGCATTTAGTCTACTGTCTATTCATTCAGAGGAGATTACGG 156802

```
QY 1529 -----AlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGln 1545
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156801 ACACACTACTGATGAGAGGAGTTCCTCTCTCAGGAAACAGTCAATTGGATTGAGAA 156742
QY 1546 LysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGln-----LeuLeu 1563
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156741 CAGTACCGTGCAAGA-----CTTCAGCGCAAGTATTGCGCAACAATTAAGTGTG 156691
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156690 GAGAGCGCGTTTACAAAGATCCAGTCATCATCAAAAGGATGATGTTAAGGAAAGTGTG 156631
QY 1584 GlnAlaLeu-----MetThrProGluMet 1591
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156630 CAGAGATGCACAGGCGCTGCTACTGATTGATTCAGGCCACTTTCGGAATGCACAGAGCTGT 156571
QY 1592 MetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyr 1611
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156570 GTGAGGTACAGCGTTTGAAGAGGGCGCTCAGTTGTGATCCAGAAAGCAATACCGAGCACAC 156511
QY 1612 LeuIleArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGln 1631
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156510 -----AGAACTGCAGAACTGCACAGCGCAGCTTTTGTTCAGGCAGACAAAGCTGCT 156460
QY 1632 LeuArgGlnGluAla-----LysAlaValAsnSerValArgCysLysVal 1646
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156459 CTGACCTTCAGGCTGCTTTAGGGCGTGAAGCGCAGAAACCAATTTGAAGACTATGCAAT 156400
QY 1647 GlnGluAla-----ValArgPheLeuArgGlyArgPheIle 1658
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156399 GCTTCTGCACACCTTATTCAGAGTAAATTTAGAGCTAGATAAGTAAAGGACAAATTCCTT 156340
QY 1659 AlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHis 1678
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156339 GCTTCTCAGAAAGAGCTGCTATATTCGTGCGAG-----AGGAAATACAGACCAACCTTTAT 156286
QY 1679 LeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAla 1698
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156285 GCCAAGCATAAATGACCAATCTTCGAGTTCAGTAAGAAAGCAGCCATTACATACAGTCC 156226
QY 1699 Ile-----ArgSerGluValAspLysGlnLeuIleGluArg 1710
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156225 TCTTTTCGAAGACTGATGCTCAAAAGAGCAGTTTACAAAGAAAGG 156184

RESULT 14
AY365047
LOCUS Pan troglodytes ASPM protein (ASPM) gene, complete cds.
DEFINITION Pan troglodytes ASPM protein (ASPM) gene, complete cds.
ACCESSION AY365047
VERSION AY365047.1 GI:37953276
KEYWORDS
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 (bases 1 to 62965)
AUTHORS Kouprina,N
TITLE Pan troglodytes genomic DNA sequence of ASPM gene
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 62965)
AUTHORS Kouprina,N.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2003) LBC, NCI, 9000 Rockville Pike, Bethesda, MD
20892, USA
FEATURES
Location/Qualifiers
1..62965
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<137..>62239
/gene="ASPM"
join(<137..433,2447..2590,2737..4216,6675..6779,
11304..11450,12956..13201,14195..14262,16489..16630,
17331..17361,17883..18058,21348..21493,21584..21669,
22199..22420,23935..24142,24229..24371,24487..24615,
28544..28738,41355..46109,50394..50560,52381..52477,
53302..53511,54505..54654,55523..55714,56176..56368,
56483..56637,58133..58309,59592..59761,62137..62239)
/gene="ASPM"
/product="ASPM protein"
join(137..433,2447..2590,2737..4216,6675..6779,
11304..11450,12956..13201,14195..14262,16489..16630,
17331..17361,17883..18058,21348..21493,21584..21669,
22199..22420,23935..24142,24229..24371,24487..24615,
28544..28738,41355..46109,50394..50560,52381..52477,
53302..53511,54505..54654,55523..55714,56176..56368,
56483..56637,58133..58309,59592..59761,62137..62239)
/gene="ASPM"
/codon_start=1
/product="ASPM protein"
/protein_id="AA04850.1"
/db_xref="GI:37953277"
/translation="WANNRVGRCWEVSPTEPPAGLRGPAABEASPPVLISHP
CRSPFLCFDVLGASRTLSLADNPNEVAEVKISHFFPADLIGSVSRQFVLQPKF
KIVISVNTPLKEGVREIMTFVNDVLKHQAILGNABEOKKKRSLWDTTKKKIS
ASTSNRVSTNTQNTKTFVSQKQVDRVSPLOACENLANNRGPTTENNLSLEBK
IPISIPAPNEYHGATCLPLSVRSITVSSLSHASENRELLNVHVASNVKSVNKEAV
TETSPSNVNGSGENSKLSITPNVSTLTNTQSTHFLSPDSPSVNNSHANNLEL
VTCLSDDPMKDNKSPVHLESTIAHEIYQKILSPDSFTKDNVGLQDLSESVNPLS
PFQELKDNWNYMCTSCQCTCKPLSNESQVPSQPEDWRKSEVSPICEQGSKPAI
FEEELVMKSNYSFIFKONNPKFSVODISSHSNKQPKRRPILSATVTKKATCFREI
QTEINPKRAKCLNSAGVEHEKVINNOKEDEFHSLPFIIDPLSKSKSYKNEVTPSI
TTASVARRKSKDSMEDGNVRVAITEHTEVREIKRIHFSPSPKTKTSVAVKTKNITPI
SKRIINREKILNKKTDLSIPRTPLSKTNKTKPIIAVAQSSLTPIPKLPTIPRHPM
PFAAKMFDYERWEKQSGFTWLNFLTDDPTVTNTI SEVNAATLLGTTGNQHKI
SVPAATKREMSLRATFARLNRRAACRLTSEKMWAKKKEIEIEARLIRVK
DRLHVKGEQKVLNLLSNPLWIRLGETTYGELISLEDNSVTGLAMFILNRL
WNPDIABYRHPVPLHYRDGHEEVLSKFTKLKLLVCLFYDAIKISLIDHDPLCF
KDAEFKASKELIAFSDRFLSGEGLSHLGLGLPVNHQVFPDFEFAVNLAVDL
QCGFRLVTMELLTQNDLSKLRIPATISRLQKMNVDIVOLVKSRLGIELSDHGT
ILSKDIDVHRREKTLLELWKTAPAFQVDLSLNDQKEIAFLKHTKSKIKTISLSC
LSDHLNKQKDCGSGFEQSEINKLMWNAVCAFYNKKNVFNTPSFSQVRLCY
LIHHYPCVFPDAICQRTQCTVECTQGVVNSSDSDSDSLMSDFSDPHENTSE
LYKELENEKNHFLVRSAGVLDGGIPAMINSHMSNTIPDEKVTIYLSFLCARLL
LRKEITRAARLIQTTWRKYLKTDLRHOEREAARIIOAVINFLAKORLKRVAAL
VIQYWRVLAQOKLLMLKLEKONKAASLIOGYWRYSTORFLKLYYSILQ
SEIPMTIATVSYRYVATVTOREHRAVLRKQDOOBYEMLSKSTLIIQSMRQWQ
RMOQSVATVILQAFREHLKQAKQNSAIIISQSWRMKEIKIIRSCVYII
QRRFQQAQKLYKKKKSILTIQKYIAYLKGKERTNYLOKRAAIIQLOAFARLLK
AHNLQIRACVIOQSYWRMRQDRFLNKKTIILQAHVRKHQORQYKRWKKAAY
IIQTHFRAYIPAMKVLASQKIRSAVILQSYARGMAQAKMIIHILTSYKIVRA
YVSKKEFLSNATIKLOSIVKMKOTRKQYLHRAAALFIOQCYRSKSLAAQKRREYV
QMBESCIKLOAFRGVLYRKOMLRKAVISLSQSYRMRKAQOYIKWYKALIVIQNY
THAYAQVNQRNKLQVKAATCLOARVGYKVRQIIKOOSIAALKIQSAFRGYNRV
KYQSVLSIIKIQRYRATLTHDTRTHFLKTAAYISQSAYRGWKVKRQRIRRHQA
ALKIQSAFRMAKQOQFRLFTTAALVIOQNFRAWTAGRQRMETLEHRAVLMQSMW
KGLTRLOQRQHKCAIIIOQSYRMHVQOKWKIMKKKALLIOKYRAYISIGREONHL
YLTKAAVTLQSAVGMKVRKRIKDCNKAATVIOQKYRAYTKKAYATVRSALIIQ
BWYRGKLTNHQHEVLANKTAKILOQSYVYIRVRHLQHMRAATFIKAMFKHQOS
RISYHTRKAAIVIOQRCAVIOQGRQREKILILKAVLQSAFRGVRVRLRQWQ
IATLIIQSNRYRYOFTYFKLKITTKVQRYRAMKERNIOQRYKRLHSHVYIIQA
IFRGKARHLKMMHIAATLIQRRFTLMRRRFLSKLTALIOKRYKRAHLYIQLL
QFLOQNAVIKIQSSYRWRMIRKMRHRAATFTQATPMRHLHMYQALQKQASVVI
QOQYQANRAAKLORHLYRORHSAVILQAPFRGMKTRRLKSMHSATLQSRFRSLI
VRRFISLKEKATIFYQRYKATIKAKHLHOPHLRKAATIOSYRYRLVWVKLQEM
QRAAVLQATFRMHRTYITFQWKGASLIIQOHTYTYRAAKLRNRYIRQWHSAVIQ
AAYGKARQLLRREKHASIIIOSTYRMYOYCFYKQLQWATKIIOEKYRANKKQKA
QRYRKLTAVRTQAVICIOQYRGFKVDKIQMHRAATLQSFYRMRKVDYETVK
FOHNLKKECTVQAGQDMNTKKQIOEQHQAIIIOQKCKAPKIRHATYVSI
TAVVIOQNYRYLVVVKTRKNFLAVQKSVRTIOAFPRGMKYOKLKNVSEKMAIV
NKSALCCTRSKTOEAVSGEWMIOEWYKASGLACQSEYHQSRAATVIOKACFRM
VTRKLETKORCAALRIOFFLOQVRRFRVQOQRAATLQHYPTWTRKQFLYLRKAA
VYLQNHRYAFISAKQOQVYTIQIRSSVYIIQARSKGTQKRFQIRKSTIKIQAHR
RYRAKYLCKVKAACKIOAWYRCWRAHEKYLILKAVKIIQCFYTKLFTFLNVR
SAIITORKWRAITLSAKIAHEHFLMKHRAACLIQAHVGYGKRGVFLRQKSAALIIQ
KYTRAREAGKREIRKIVIEFKKSTVILQALVRGLVYRKRILEQRAKIRLLHFTTAAAYH
```

LNARVIAQAYKLYLAVKNANKOVNSVICIQWFRARLOEKRFIOKVHSIKKIBEGOE  
 CLSORNEAAVIOKAVHFVLRKOKKFTSIIKIQALWGYSWRKNDCTKIKARL  
 SLOVNRREIRNKLYRTALALHYLLTYKHLNLSALEKHLVFRRLSPICENWAO  
 SGALSIFLNRNSIPCMVEIRYAVQVLLNVSKEYKTSIAVDIVENCIDILELLEU  
 QIYREKPNVADKGSIFTKTCLLAILKTTNRASDVRSKVDRIYSLYKLTAAH  
 KHWNTIRILYKOKNSSISIPETIPFVTRIVSRILKPDVLRDNMBEITNPLOAI  
 QWMDTLGIEY"

## ORIGIN

## Alignment Scores:

Pred. No.: 7,87e-31 Length: 62965  
 Score: 744.50 Matches: 286  
 Percent Similarity: 41.29% Conservative: 207  
 Best Local Similarity: 23.95% Mismatches: 379  
 Query Match: 7.88% Indels: 322  
 DB: 9 Gaps: 43

US-09-914-698-1 (1-1861) x AY365047 (1-62965)

Qy 868 TrpGlnLeuTleTyrLysPheArgSerProLysPheHisAlaAlaLthrValLeuGln 887  
 Db 41655 TGGCATTTA-----AGAAACACAGCTAAAGAACAAATCTCTGTTATATACATACAA 41705  
 Qy 888 LysTrpTrpArgArgHis-----TrpLeuHis-----Val 897  
 Db 41706 TCATGTTATGAATGCATAAAGATTACGGAAATATATATTATATTAGATCTTGTGTGTT 41765  
 Qy 898 ValIleGlnArgArgIleArg-----HisLysGluLeuMetArgArgHisArg 913  
 Db 41766 ATCATTCAGAAAGATTTCGGTGTCTTCAAGCCCAAAAGTTGTATAAAGAAAGAAAGAG 41825  
 Qy 914 AlaAlaLthrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeu 933  
 Db 41826 TCCATACTAACCATCCAGAGTACTACAAGCATATCTGAAGAAAGATTGACGGCACC 41885  
 Qy 934 PheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeu 953  
 Db 41886 AACTATTTCAGAAACAGAGCTGCAGCCATCAATTACAAGCTGCTTTTAGGAGA---CTG 41942  
 Qy 954 AlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArg 973  
 Db 41943 AAAGCTCAATATTATATAGACAAATTAGAGCTGCTTGTGTATTATTCAGTCACTAGG 42002  
 Qy 974 AlaGlnGln----- 976  
 Db 42003 ATGAGACACACAGAGTTCGATTTTAAACCTTAAGAGACTATTATCAAAATTCACGCA 42062  
 Qy 977 -----LeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePhe 994  
 Db 42063 CATGTAAGAAACATCAACACGACAGCAAAATATAAAGATGAAGAAAGCAGCTGTTATA 42122  
 Qy 995 LeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaAlaGluThr 1014  
 Db 42123 ATTCAGACTCAATTCGAGCTTATATTGTCATGAAAGTTCTAGCATCTTACACAGAAA 42182  
 Qy 1015 AlaArg-----LeuGlnArgSerGlnLys-----GlnGlnAlaAlaAlaSer 1028  
 Db 42183 ATAGGCTCTGCTGTCATTTGCTGCTGCTGCATATAGAGGGATGCAAGCCAGGAAATG 42242  
 Qy 1029 TyrIleGln-----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGln 1043  
 Db 42243 TATATTCAATCTCATCTGTTTATAAAGATTCAATCAVATATTATCGTGTATTGTTTCT 42302  
 Qy 1044 ArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSer 1063  
 Db 42303 AAAAGGAATTTTGGACCTAAATAATGCTACATA-----AAATTCAGATCA 42350  
 Qy 1064 LysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsn 1083  
 Db 42351 ATTGTAAAGTGAACAAACAGTAAACAATATTGTCATTTAAGACAGCTGCTACTATT 42410  
 Qy 1084 IleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAla 1103  
 Db 42410

Db 42411 ATCCAGCATGTTTACCCTTCCAAAAAATAGTCGACAAAAAGAGAGAGTATATGCAG 42470  
 Qy 1104 LeuArgSerValLeuLysValGlnAlaTyr----- 1114  
 Db 42471 ATGCGGAATCTTGTATCAAAAGTCGAAGCATTTGTTAGAGATACCTTGTCCGAAAGCAG 42530  
 Qy 1115 -----ArgLysAlaThrIle-----GlnMetArgIleAsp 1124  
 Db 42531 ATGAGGTTACGAAGAAAGCTGTTATTTCACTACAGTCTTATTTCAGAAAGAGGCT 42590  
 Qy 1125 ArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAla 1144  
 Db 42591 CAGCAGTATTATCTGAAAAATGTATAAAGCAATTATTGCTTTCAGAAATTAATCATGCA 42650  
 Qy 1145 IleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeu 1164  
 Db 42651 TACAAAGCACAGGTCATATCAGAGGAAGAACTTCTTGAAGTCAAAAAGCAGCTACTTGC 42710  
 Qy 1165 ValGlnLysArgTyrArg-----MetArgGlnGlnMetIle--- 1176  
 Db 42711 TTGCAAGCAGCTTACAGAGGTTATAAAGTACGCCAGCTAATCAAAACAACAATCTATAGCT 42770  
 Qy 1176 -----ArgLysAlaThrIle----- 1176  
 Db 42771 GCTCTTAAATTCAGTCTGCTTTTAGAGGCTATAATAAAGGTAAAAATATCAATCTGTG 42830  
 Qy 1177 -----GlnAsp 1178  
 Db 42831 CTTCAATCTATATAAAGATTACAGAGTGTGTACAGGCGGTACAGACTCTTCATGATACA 42890  
 Qy 1179 ArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGln-----ArgArg 1195  
 Db 42891 AGAACATATTTTGAAGACAAAGGCGCTGTGATTTCCCTCCAGTCTGCTTATCGTGGC 42950  
 Qy 1196 TrpArgAlaThrLeuGlnMetArgGluArgLysAsnTyrLeuHisLeuGlnThrThr 1215  
 Db 42951 TGGAAAGTTCGAAACACAGATTAGAAGGGAACATCAAGCTGCCTTG----- 42995  
 Qy 1216 ThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu 1235  
 Db 42996 -----AAGATTCACTGCTGCTTTTAGA-----ATGGCCAAAGCCGCAAGAACRG 43037  
 Qy 1236 PheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGln 1255  
 Db 43038 TTTAGATTGTTTAAACACAGCAGCATTAGTCAATCCGCAAAATTTTCAGAGCATGAGACTCA 43097  
 Qy 1256 MetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGln--- 1274  
 Db 43098 GGAAGGAAGCAACGATATGGAGTATATTGAATCCGTCATCGGCTACTGATGCTTCAATCT 43157  
 Qy 1275 -----ArgArgPheHisAlaGlnLysSer----- 1282  
 Db 43158 ATGTGGAAGGAAAAACACCTGAGAAAGACAGCTTCAAAGGCAACATAATATGTGCTATCATC 43217  
 Qy 1283 -----MetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAla 1296  
 Db 43218 ATACAGTCATATATAGATGTCATGCAACAAAGAAAGTGAATAATCATGAAAAAGCT 43277  
 Qy 1297 ValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSer 1316  
 Db 43278 GCTCTTCTGATTCAAAAGTATTATAGGCTTATAGGCTTATGGAAGAGACAGAAATCATTTA 43337  
 Qy 1317 PheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsn 1336  
 Db 43338 TATTGAAAAACAAAGCAGCTGTGTTAACTTTTACAGTCAGCTATCTGTGT----- 43388  
 Qy 1337 MetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThr 1356  
 Db 43389 ---ATGAAAGTGAAGAAAGAAATAAAGATTGCAACAAAGCAGCAGTCACTATACAGTCT 43445  
 Qy 1357 ArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGlu 1376  
 Db 43446 AAATACAGAGCTTACAAAAACCAAAAGAAA-----TATGCAACCTATAGAGCT 43493



```

/tissue type="human esophagus tumor"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host
DH10B_sites SfiIA + SfiIB"
/dev_stage="adult"
polyA_signal      5718..5723
polyA_site        5738
ORIGIN

```

```

Alignment Scores:
Pred. No.:      8.49e-32      Length:      5760
Score:          737.00      Matches:      408
Percent Similarity: 38.35%      Conservative: 331
Best Local Similarity: 21.17%      Mismatches: 637
Query Match:      7.75%      Indels:      554
DB:              9      Gaps:      77

US-09-914-698-1 (1-1861) x HSM808955 (1-5760)
Qy 253 HisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProArgThrPro 272
Db 451 CAITCAGAATTACTATCATGCATA---CAAAGCACAGTCAATCAGAGGA-AGAACTTCT 506
Qy 273 ThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArg 292
Db 507 TGCAAGTCA-----AAAAGCAGCTACT-TGCTTGCAAGCAGCTTACAGA 550
Qy 293 LysTyrSer-----GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyr 310
Db 551 GGTATTAAGATCCAGCTCAATCAACAACAACTATAGCTGCTCTTAAATTCAGTCT 610
Qy 311 AlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsn 330
Db 611 GCTTTAGAGGCTATAAT-----AAAAGGTAATAATATCAATCTGTGCTTCAATCT 661
Qy 331 -----ArgTyrTyrGlnGlnGlnGlnIleGlnIleLysAlaLysAspLeu 345
Db 662 ATAATAAAGATTCAGAGATGGTACAGGCGTAC-----AAGACTCTT 703
Qy 346 AsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlnGluPheLeuPheAsnHis 365
Db 704 CATGATACA-----AGAACACATTTTGTGAACACAAAG 736
Qy 366 SerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGlyArgLysSerVal 385
Db 737 GCAGCTGTGATTCCCTCCAGTCTGCTATCGTGGCTGGAAGGTT---CGGAACACAGATT 793
Qy 386 LysGlySerProValLysAsnProHisLysArgArgSerHisGluLeuSer----- 402
Db 794 -----AGAAAGGGAACATCAAGCTGCCTTGAAGATT 823
Qy 403 -----PheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThr 417
Db 824 CAGTCTGCTTTAGATGGCCAGCCAGAACACAGTTTAGATTGTTTAAACACAGCAGCA 883
Qy 418 ValAlaIle-----SerProLysLysLysGlnArgValGlu--- 429
Db 884 TTAGTCATCCAGCAAAATTTTCAGAGCATGAGTGCAGGAAGGAAGCAATGATGAGATAT 943
Qy 430 -----AspThrThrLeuPro 434
Db 944 ATTGAATCCGCTCATCGCGTACTGTGCTTCAATCTATGTGGAAGGGGAAACATCAGTA 1003
Qy 435 ArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTyrProHis 454
Db 1004 AGACAGCTTCAAAGGCAACATAAATGTCTATCATCATACAGTCATCATATAGATGCAT 1063
Qy 455 AlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThr 474
Db 1064 GTGCAACAAAGAAAGTGGAAA-----ATCATCAAAAAGCTGCT--- 1102
Qy 475 ProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyr----- 490
Db 1103 -----CTTCTGATTCAAAGATATATATAGGGCTTACAGATTGGGAAGA 1144

```

```

Qy 491 ---AspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThr 509
Db 1145 GAACAGAAATCATTTATATTGAAACA-----AAAGCAGCTGTAGTA 1186
Qy 510 ThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529
Db 1187 ACTTTA----- 1192
Qy 530 GlnAlaAspPheLysLysTyrLeuAsnAlaLeuValSerIleProAlaAspLeuAspAla 549
Db 1193 CAGCAGCTTATCGT----- 1207
Qy 550 AspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeu 569
Db 1208 -----GGTATGAAAGTGAGAAAAGAAATAAAGGATTGC---AACAAAGCAGCA 1252
Qy 570 ValValAlaProThrLysGluGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGlu 589
Db 1253 GTCACATATACAGTCTAAATACAGAGCT-----TACAAAACCAAAAAGAAATATGCA 1303
Qy 590 ThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSer 609
Db 1304 ACCTATAGAGCTTCAGCTATATAATT-----CAGAGATGGTATCGAGGTATT 1351
Qy 610 LysValAlaValTyrValAsnLysGlnAlaLeuArgIle-ArgSerAspArg----- 626
Db 1352 AAAATTACAAACCATCAGCATAGGAGTATCTTAAATTGAAGACAGACGCAATTAAAAATC 1411
Qy 627 -AsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeuCysPheAs 646
Db 1412 CAATCTCTTTATAGAGTA-----TTAGAGTTAGAAGACATATTCAA 1453
Qy 646 n-----ProLeuTyrLeuArgLeuGluValValPheGlyGluLy 661
Db 1454 CACATGCACAGGGCAGCCACTTTTATTAAAGCCATGTTTAAAGTCATCATCAGCAATA 1513
Qy 661 sIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLe 681
Db 1514 AGTTACCATACATGAGAAAAGCAGCTATTG----- 1546
Qy 681 uPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTy 701
Db 1547 ATTCAAAGTA-AGATGT-----AGAGCATATTATCAGGTTAAATGACGCGTGAAGATA 1599
Qy 701 rAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAs 721
Db 1600 CCTCACAATTTGAAAGCTGTTAAAGTCCTTCAGGCAAGTTT----- 1642
Qy 721 pGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerPr 741
Db 1643 -----AGAGGAGTAAGA-----GTTAGACGCACTCT 1668
Qy 741 oHisLys---GluThrLysAspIleLeuLeu-----ArgPheSerSerG1 755
Db 1669 TAGAAGATGACGAGCTGACGACACACACTCATTTCAGTCAAACTACAGAGATACAGACAGCA 1728
Qy 755 uLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuG1 775
Db 1729 AACATACATTAT-----AAGTTAAAGAAATAACAAAACAGATACA 1770
Qy 775 nHisArgGlnThrPheLeuAspGluPheAspTyrAlaPhe-----AsnAsnLeuAl 792
Db 1771 GCAAGATATCTGGGCAATGAAAGAAAGAAACATACATAATTTCAAAGGTATACAACTGAG 1830
Qy 792 aValAspLeu-----ArgAspGlyValAr 800
Db 1831 GCATTCTGTATATACATTCAGGCTATTTTAGGGGAAAGAAAGCTAGAGACATTAAA 1890
Qy 800 gLeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVa 820
Db 1891 AATGATGCATATACCCGCAACTCTCAITTCAGAGGAGATTTAGAACTCTAATGATGAGAAG 1950

```







Db 3985 CTGCAAGATTCAAGCCTGGTATAGATGTTGGAGAGCACACAAAGAAATATCTAGCTATATT 4044  
QY 1426 ----- 1426  
Db 4045 AAAAGCTGTTAAATTAATTCAGAGTTCCTTCTATACCAAACTAGAGAGAACACAGTTTTT 4104  
QY 1427 -ArgileLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaAr 1446  
Db 4105 GAATGTGAGAGCATCAGCAATATATCATTCAGAGAAATGAGAGCTATACTTCTCTGCAAA 4164  
QY 1446 gArgGlnArgGlnIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheII 1466  
Db 4165 GATAGCTCATGAACACATCTCTTAATGATAAAAGA----- 4198  
QY 1466 eaArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgAlaAl 1486  
Db 4199 -----CATCGAGCTGC 4209  
QY 1486 aValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTy 1506  
Db 4210 TTGTTTTCATCCAAAGCATTATAGAGCA-----TATAAGGAAGCAGGTCIT 4257  
QY 1506 rGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526  
Db 4258 TCTTCGGCAGAACTCTGCTGCTTTGATCATACAAAATATATACGAGCCAGGGAGGCTGG 4317  
QY 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLy 1546  
Db 4318 AAAGCATGAAGGATATAATATATGTAATTTAAANAATCTACAGTTATCTCTACAGCACT 4377  
QY 1546 sPheArgGly-----LysArgLeuMetIleGlnArgAsnCysPheGlnLeuLe 1563  
Db 4378 GGTGGGTGGTGGCTAGTACGAAAAGATTTTATAGACAGAGAGCCAAATTCGACTTCT 4437  
QY 1563 uArg----- 1564  
Db 4438 TCACCTTCACTGCAGCTGCATATTATACCTCAATGCTGTGTAGAAATTCAAAAGAGCCTATAA 4497  
QY 1565 -----CysSerMetPr 1568  
Db 4498 ACTTTACCTGGCTGTGAAGAAATGCTAAACAGAGGTTAATTCAGTCACTGTATTCAGAG 4557  
QY 1568 oGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe---GlnAlaLeuMe 1587  
Db 4558 ATGGTTTCGAGCAAGA-----TTACAAGAAAAGAGATTTATTCAGAAATATCA 4605  
QY 1587 tThrProGluMetMetAsp-----LeuIleArgGlnLysArgAlaAl 1601  
Db 4606 TAGCATCAAAAAGATTGAGCATGAAGGTCAAGAAATGCTCGAGCCAGCGAAATAGGGCTGC 4665  
QY 1601 aLysValIleGlnArgTyTrpArgGlyTyTrpLeuIleArgArgGlnLysHis 1619  
Db 4666 ATCAGTAATACAAAGACAGCATGGCCATTTTCTCCCTCGTAAAGAACGAGAAATTCAC 4725  
QY 1620 -GinglyLeuLeuAspIleArg----- 1626  
Db 4726 TAGTGAATCATTTAAATTCAGGCATTATGAGAGGCTATTCTTGGAGGAGAAATAATGA 4785  
QY 1627 ----LysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLy 1645  
Db 4786 TTGTAACAAAATTAAGCTATACGACTAAGTCTTCAAGTTGTTTAATAGG-----GA 4836  
QY 1645 sValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaVa 1665  
Db 4837 GATTCGAGAGAAAACAAACTTACAAAAGAACTGCACCTTGCACTTCATTACCTTTTGAC 4896  
QY 1665 lLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGI 1685  
Db 4897 ATATAGCACCTTCTGCCATTCTTGAGGCTTTAAACACCTAGAGGAGTAGTTACTAGA-- 4954  
QY 1685 uPheMetSerThrPheCysTyTrpGlyIleMetAlaGlnAlaIleArgSerGluValAspLy 1705  
Db 4955 ----TTGTCTCCACTTTGTTGTGAGAAACATGGCCAGAGTGGAGCAATTTCTAAAAATATT 5010

QY 1705 sGlnLeuIleGluArgCysSerArg----- 1713  
Db 5011 TGTTTTCATCCGAAGTTGTAATCGCGTATTCTTGTATGAAGACTATCGAGATGCTGT 5070  
QY 1714 ----IleIleLeuAsnLeuAlaArgTyTrpAsnSerThrThrValAsnThrPheGlnGluGI 1732  
Db 5071 GCAAGTCTTCTTAATGTAATCTAAGTATGAGAAACTTACTTTCAGCAGCTTTATGATGAGA 5130  
QY 1732 yGlyLeuValThrIleAlaGlnMetLeuLeu-----Ar 1743  
Db 5131 AAATTCGTATAGATATACTATTGGAGCTTTTGCAGATATATACCGAGAAAAGCCTGGTAATA 5190  
QY 1743 gTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAl 1763  
Db 5191 AGTTTCAGACAAAGCGCGAAGCATTTTACAAAACCTTGTGTGTGTGCTATTTACT 5250  
QY 1763 aHisCysProLysLysArgLysIleIleHisAspTyTrpMetThrAsnProGluAlaIleTy 1783  
Db 5251 G-----AAGACAACAATAGAGCCTCTGTATGTACGAAGTAGGTCCAAAGTTGTTGA 5301  
QY 1783 rMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaAr 1803  
Db 5302 CCCTATTTCAGCTCCCTACAACTTACAGCTCTATAACATAAAATG-----AATACTGA 5355  
QY 1803 gLysProProProMetThrSerGlyArgTyLysSerGlnLys-----II 1818  
Db 5356 AAGAATA-----CTTTACAAGCAAAAAGAGAATTCCTTCTATAAGCAT 5397  
QY 1818 eAsnPheThrProCysSer-----LeuProSerLeuGluProAspPh 1832  
Db 5398 TCCTTTTATCCAGAAAACACTCTGAAGGACCAGAAATAGTTTCAAGACTTAAGCCAGATTG 5457  
QY 1832 eGlyIleIleArgTySerProTyThrPheIleSerSerValTyTrpAlaPheAspThrII 1852  
Db 5458 GGTTTTGAAGAGAGATAACATGGAAGAAATACAAAATCCCTCGAAGCTATTCAAATGGT 5517  
QY 1852 eLeuCysLysLeuGlnIle 1858  
Db 5518 GATGATACGCTTGGCATT 5536

Search completed: September 15, 2004, 06:57:05  
Job time : 17563 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: September 14, 2004, 21:44:50 ; Search time 1427 Seconds  
(without alignments)  
5540.222 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPTO\_spool\_h/US0914698/runat\_10092004\_150004\_28194/app\_query.fasta\_1.2055  
-DB=N Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0914698 @Cgn\_1\_1091 @runat\_10092004\_150004\_28194 -NCFU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002s: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004s: \*

Print. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9514	100.0	5586	3 AAA37761	Aaa37761 Drosophil
2	9482	99.7	6003	4 ABL11881	Ab111881 Drosophil
3	9386.5	98.7	8330	4 ABL11880	Ab111880 Drosophil
4	834	8.8	2453	4 AAH14237	Aah14237 Human cDN
5	707.5	7.4	3887	4 AAH14236	Aah14236 Human cDN
6	700.5	7.4	4027	6 ABQ99318	Abq99318 Human cod
7	542.5	5.7	1115	5 ABV20287	Abv20287 Human pro
8	542.5	5.7	1115	5 ABV20247	Abv20247 Human pro

C 9	542.5	5.7	1115	5	ABV26078	Human pro
C 10	542.5	5.7	1115	5	ABV20220	Human pro
C 11	542.5	5.7	1115	5	ABV26049	Human pro
C 12	542.5	5.7	1115	5	ABV26118	Human pro
13	439.5	4.6	752	5	ABV09513	Human pro
14	437	4.6	2414	4	AAH14270	Human cDN
15	437	4.6	2414	7	ACC50994	Human bla
16	353	3.7	576	5	ABV30689	Human pro
17	353	3.7	576	5	ABV30689	Human pro
18	342	3.6	24789	5	ABV39657	Human pro
C 19	332	3.5	802	5	ABV25946	Drosophil
C 20	332	3.5	802	5	ABV25946	Human pro
C 21	332	3.4	7677	5	ABV20113	Human pro
22	322	3.4	7677	5	ABV23205	Human pro
23	322	3.4	14800	6	ABV29042	Human pro
24	317.5	3.3	15231	6	ABV66291	lung canc
25	317.5	3.3	15231	7	ABR63598	Abk63598 Rat seque
26	317.5	3.3	15231	9	ABT42078	Toxicity
27	317.5	3.3	15231	9	ADB59205	Toxicity
28	316.5	3.3	14835	6	ADB53026	Primary r
29	316	3.3	771	4	AA594858	Human DNA
30	316	3.3	771	4	AAH05820	Human cDN
31	303.5	3.2	8673	4	ACC50236	Breast ca
32	300.5	3.2	476	5	ABL11109	Ab111109 Drosophil
33	300.5	3.2	476	5	ABV39583	Human pro
34	300.5	3.2	476	5	ABV30748	Human pro
35	300.5	3.2	476	5	ABV39716	Human pro
36	294	3.1	9551	2	ABV09572	Human pro
37	293	3.1	5229	8	AAZ22301	cDNA enco
C 38	286	3.0	11200	4	ABH83244	Human cDN
39	285.5	3.0	501	5	ABV09438	Human pro
40	285.5	3.0	5928	4	ABL10555	Drosophil
41	282.5	3.0	6946	5	ABV25339	Human pro
42	279.5	2.9	461	5	ABV00403	Human pro
C 43	279.5	2.9	8527	4	AAI60039	Human pol
44	278.5	2.9	5230	7	AAI60039	Human ves
45	277	2.9	5068	5	AAF32508	Human mal

## ALIGNMENTS

RESULT 1

AAA37761

ID AAA37761 standard; DNA; 5586 BP.

XX AAA37761;

AC AAA37761;

DT 04-DEC-2000 (first entry)

XX Drosophila Asp coding sequence.

DE Drosophila Asp coding sequence.

XX Drosophila Asp coding sequence.

KW Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor;

KW tumour cell; ds.

XX Drosophila sp.

OS Drosophila sp.

XX Drosophila sp.

PH Key

FT CDS

FT 1..5586

FT /\*tag= a

FT /product= "Asp"

XX WO200052478-A1.

XX 08-SEP-2000.

PD 03-MAR-2000; 2000WO-GB000785.

XX 04-MAR-1999; 99GB-00005007.

PR (UYDU-) UNIV DUNDEE.

XX Glover DM, Avides MDC;

XX

DR WPI; 2000-594203/56.  
DR P-PSDB; AAY90350.  
XX  
PT Use of Drosophila Asp polypeptide for identifying substances capable of  
PT disrupting microtubule organization center integrity and use of the  
PT identified substances for inhibiting mitosis in tumor cell.  
XX  
PS Disclosure; Page 44-47; 51pp; English.  
XX  
CC This sequence encodes the Drosophila Asp protein. The invention relates  
CC to the use of Drosophila Asp polypeptide (or its homologue, or fragment)  
CC capable of stimulating formation and/or maintenance of microtubule  
CC organising centres (MTOCs), in an assay for identifying a substance  
CC capable of disrupting MTOC integrity. Asp polypeptide or its homolog is  
CC useful for identifying a substance capable of disrupting MTOC integrity.  
CC Substances identified by the method can be used to inhibit mitosis, e.g.  
CC in tumour cells  
XX  
SQ Sequence 5586 BP; 1532 A; 1466 C; 1387 G; 1201 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 0 Length: 5586  
Score: 9514.00 Matches: 1861  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0  
US-09-914-698-1 (1-1861) x AAA37761 (1-5586)  
QY 1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuLe 20  
DB 1 ATGGAGCTAGTGTGGAGCCCGCTTGTAGTGCGCTCGAAGGAGACGCTGCAGCTAATA 60  
QY 21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40  
DB 61 GACAAACCGCAACTCCGAAAGAGGTGATGATCATCTCAAGTCCAAAGACCAACGACCG 120  
QY 41 ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60  
DB 121 GTCAAGNACCGCGCAAAATTCCTACTGTGCGNAGACCTGCGAGCTGAATTCGCCGACA 180  
QY 61 GlyAlaGlyLysThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMet 80  
DB 181 GGAGCTGGCAAGCAATGAAAGCGTGTATCGCTCTGTGCGAGCAAAAAGAGCGCATG 240  
QY 81 SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro 100  
DB 241 TCTGCGACGACGCGCTCCCTCCAGCAGACATGGCGAGTGACTGTCTCTCGCGTCCC 300  
QY 101 AlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLysThr 120  
DB 301 CTTGCGATGGGCATATCCACTCCACAGGCTCTCTTGTGAGAGAGAAATGATATACAGACT 360  
QY 121 ProGlnGluGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer 140  
DB 361 CCACAAAGAGCGCGTATACATATATCCACAGCCTCGCAGCTCTTAAGGAAAAATCTAAGC 420  
QY 141 ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThrProLeuThr 160  
DB 421 CCTATGACGCGAGAAACCTACTCGAGCTGATTGACAAATCTGGCAATTCACACCTCTCACC 480  
QY 161 GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr 180  
DB 481 GAAACCCGTGGCAAGAGACAGCTACCATTTTCCCGGACAAATCTGGCAGCGCTGGCCACA 540  
QY 181 ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgIleThr 200  
DB 541 CCAACACTTAAGGGAATGAAATTCATGTCTGAATGATATGCGCGCGCTGCAATCACT 600  
QY 201 ProAspAsnLeuGluAsnGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu 220  
DB 601 CCAGATGATCTAGAAGATCAGCCCTGCCACAAACAAACGTTTCGATGTAAAGCATTCGAG 660  
QY 221 ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr 240  
DB 661 ACTATCAATATTCGTGGACACCTTGGACTGCTCCAGGATCGATGGACAAACCCGATACG 720  
QY 241 ProLeuAsnLysThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle 260  
DB 721 CCCCTAAATAAGACAACAACCATTTGTCATGCCACGCACACACAGAGCTCTGGCGTGTATT 780  
QY 261 HisGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280  
DB 781 CATGAGGAGGAGGAGCAAGTCCACCTAGGACCCCAAGAGAGCGCATACAGACCTG 840  
QY 281 LysArgAspIleLysLeuValGlySerProLeuArgLysTyrSerGluSerMetLysAsp 300  
DB 841 AAGAGGGACATTAAGTTGGTGGGTTCCACCTTACGAAAGATTTCCGAGTCCATGAAAGAT 900  
QY 301 LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnLysSerMetProAsnLeuAsn 320  
DB 901 TTGTCTCACTTCTATCGCCACAACCTAAGATATGCCATTCAAGGCTCTATGCCCTAATCTAAAT 960  
QY 321 GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlnGlnIleGlnIle 340  
DB 961 GAAATGAAATCCGCTCGATCGAACAGATAGATATTTACCAGGAGCAGAGATCCAGATA 1020  
QY 341 LysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu 360  
DB 1021 AAGGCCAAGACTTGAATAGTCTCTAGTAGGAGGCTAGTTTGGCGCGCAGCAGAG 1080  
QY 361 PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluVal 380  
DB 1081 TTTCTATTCAACCACACATGAGATCTCGCTCAGTCCAGTCCGTTTAACTCTCCATGAAGTA 1140  
QY 381 GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu 400  
DB 1141 GGTGGAAGTCGGTGAAGGAAGTCCGGTGAAGATCTCCACAGGCCCTCTCTCATGAG 1200  
QY 401 LeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThrValAlaIle 420  
DB 1201 TTGAGTTTTTCGAGTGCACCTAGCAACGAATCATTTGACCGCAATGAACTGTAGCCATT 1260  
QY 421 SerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla 440  
DB 1261 TCCCTCTCTAAAAAGCAACCGGTTGAGGACACTACTCTGCCCAAGTGCAGCGCCGCA 1320  
QY 441 AsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPhe 460  
DB 1321 AATGCATCTCAAGAGACGATGATGCCACGCTGGCCACACGCCCAATCCAGAAAGTTT 1380  
QY 461 LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp 480  
DB 1381 AAGCTAGCACAAACCATGTCACTGATGAGAGAGCCGCCACACCCAGAAAGTCAGGGAC 1440  
QY 481 ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle 500  
DB 1441 ACTAGCATTCAGCGCTTCGFTCAAGCTCTATGACTCGGAGCTGTATATGCGAGCGTGATC 1500  
QY 501 AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr 520  
DB 1501 AACCCGATCCATTTGCGCAACTACGACAAATTTGATCCATTTCTGGCATCTACCATGTAT 1560  
QY 521 LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysThrLeuAsnAlaLeu 540  
DB 1561 TTGGATGAACAGCGCTGGGATCGTCAAGCTGACTTTTAAAAAGTGGTAAAAATGCCCTT 1620  
QY 541 ValSerIleProAlaAspPheAspAlaAspLeuAsnLysIleAspValGlyLysLeu 560  
DB 1621 GTCTCCATACCCGCTGACCTGGACGCGAGATTAAATAACAAAATAGACGTTGGTAAGCTG 1680  
QY 561 PheAsnGluValArgAsnLysGluLeuValValAlaProThrLysGluGlnSerMet 580  
DB 1681 TTTAACGAGTGGCCACAAAGAGCTCGTGTGGTCTCCACCAAGGAGGAGCAGCTCTATG 1740

581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600  
 1741 AACTACCTAACGAATACCGCTGGAGAGCTTGGTAAGCGGCTGGAGCTCTCTTC 1800  
 601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
 1801 AGTGAGCAGATCGCGTGCATGCTCAAAAGTGGCGTATATGTCAACAAAGCAAGCTCTG 1860  
 621 ArgGileArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
 1861 CGATCCGAGGATCGTAACTTCCACCTAGACGATGATGATGCAACGACCATCTAG 1920  
 641 LeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlu 660  
 1921 CTGCTGCTTCTCAATCCCTTTGGCTGGCTTGGACTGGAAGTGTCTTTGGCGAG 1980  
 661 LysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
 1981 AAGATCCAGATCGAGTCTAATCGAGACATTTAGGCGCTCAGCACCTTTATCTCAATCG 2040  
 681 LeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
 2041 TTGTTCCGGAATAGGTGAGGAGCAGAGGTACAGCAAGCGCATACACTCCACGAGAG 2100  
 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720  
 2101 TACGCGGAGACCATTAAGAGCACTCATTTGCAGAAAATCTCTTTTGTCTTCTCTC 2160  
 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
 2161 GATCAAGCTTAAGCAGAGCGCATCGTCAAGCAATCCCTGTTTGTGTTTAAAAAGTGG 2220  
 741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerLeuLeuLeuAlaAsnIle 760  
 2221 CCACATAAGAGACCAAGGATATCTGCTGCCCTTCTGCTGGAGCTGCTCGCCCAACATT 2280  
 761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
 2281 GGTGATATTACGCGGAACTTCTGCTGCCCTGGCTACGTTCTACAGCACCGCCAAACATT 2340  
 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
 2341 TTGAGCAGAGTTCAATATAGCTTCAACAACTTGGCTGTGGCTTAAGAGATGGCGTGA 2400  
 801 LeuThrArgValValGluValIleLeuLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
 2401 CTACCCGAGTCTGGAGGATATTTTACTACCGGATGATCTAACCCGCGCATTAAGGGTG 2460  
 821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
 2461 CCAGCCATCTCTCGCTTCAGCGGATCTCAATGTAAAGCTAGCTCTGGCGCACCTTGGT 2520  
 841 GluAlaAsnPheGlnLeuGlyIleAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
 2521 GAAGCAACTCTCCAGCTAGCGCGCACATCGCGCCCAAGACATCGTTCAAGGACATCGT 2580  
 861 GluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
 2581 GAGAGAGCGCTTCTCGCTCTGGCAACTATTTTACAAATTCGCTCGCCCAAGTTTCAT 2640  
 881 AlaAlaThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGln 900  
 2641 GCGCGGCCACCGGTCTCCAGAAATGGTGGCCGCTCACTGCTGCAGCGTTGTATCCAG 2700  
 901 ArgArgIleArgHisLysGlnLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
 2701 CGTCGATTCGCGACAAAGAGCTTATGGTTCGCCACCGGCGCTACTGTCTCATTCAGGCC 2760  
 921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
 2761 GGTTCGCGGCCACACAGATGAGAAGTACGTGAAGTTGTTTAAACGCGAACGCACTCAG 2820  
 941 AlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960

2821 GCCCGCAATAATTCTGCAAAAGTTCCACCGTCCGTTATTTGGCCCGAGAGCAGTATATCAG 2880  
 961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
 2881 AGCTATCATCAGTATTATCACCATCATCGCGCTGGTGGCGAGCCCAACAACTGGGAAGGCGAG 2940  
 981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheIleGlnArgIleTyrArg 1000  
 2941 CACCCCGCAGCGGTTTGTGAGCTCCGAGAGGCTGCGATCTTTCTCAGCGGATCTGCGCG 3000  
 1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
 3001 CGACGACTCTTTGCCAAAAAACTATTGGCGGGCGGGAACAGCCAGACTTCCAGCGATCG 3060  
 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGly 1040  
 3061 CAAAAACAACAGCGAGCTGCTAGTTATATTTCAATGCAATGGCAACGTTATCAGCTGGGC 3120  
 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
 3121 AGAATTCAGCGACACAGGTTCTCGCCGAGAGGACCTCATCATCTTTGTTTCAGCGCGAG 3180  
 1061 MetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
 3181 ATGCGAGCAAGTGGAGTATGCTGGAGCAGCGCAAGAGTTCCAGCAACTAAAGCGTGCA 3240  
 1081 AlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
 3241 GCTATAAATATCCAAACACGCTGGCGAGCGAAGCTTTCAATGAGAAAGTGCAACGCTGAT 3300  
 1101 TyrLeuAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
 3301 TATTTGGCACCTTCTGTTCCAGCGCTTTAAAGTTACAGGCTTACAGAAAGCCCAATCCAG 3360  
 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
 3361 ATGAGATAGATCGTAATCAGTACTATTTCCTGCGAAAAAATGTTATCTGCTCTGCAACAG 3420  
 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
 3421 CGACTGAGGCGCATCATGAAAATGCGCAACAGAGCGAAAAATTAATCTGAGGCTGCGAAAT 3480  
 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
 3481 GCTTCAATACATAGTTCAAAAACGCTACCGCATGCTCAACAAATGATCCAGATGAAAT 3540  
 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
 3541 GCATATTTAAGAACCCGCAATGATATCATCAATGTTCCAGAGCGCTGGAGAGCCACTCTG 3600  
 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
 3601 CAGATGCTCGAGAAAGGAAGAACTTACCTTCAATCTCCAAACGACAAACAAACGAATCAA 3660  
 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheIleGlnLys 1240  
 3661 ATCAAGTTCCGTGCGCAGCGGAATGAAAGCAAGAGCGAGTTCTTCTCAGCTGAAA 3720  
 1241 LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 3721 AAGTTACCTTGTAGTCCAGAAAGCTCGCGGGCTTTGCTGCAGATCGGAAAGAGAGCGC 3780  
 1261 GlnGluTyrLeuHisLeuArgGluValThrLysLysLeuGlnArgPheHisAlaGln 1280  
 3781 CAGAGTACCTTACATCTGCGGAGTTTACCATAAAGCTGCGAGCGAGATTTCATCTCAA 3840  
 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
 3841 AAATCAATGCGGTTTCATCGAGCCCAAGTACCGCGCACCCAGGCTGCTGTGAGCTGCGCTG 3900  
 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320



PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 XX interactions.

PS Claim 1; SEQ ID NO 30125; 21pp + Sequence Listing; English.

XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 6003 BP; 1629 A; 1566 C; 1502 G; 1306 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 6003  
 Score: 9482.00 Matches: 1855  
 Percent Similarity: 99.89% Conservatives: 4  
 Best Local Similarity: 99.68% Mismatches: 2  
 Query Match: 99.66% Indels: 0  
 DB: 4 Gaps: 0

US-09-914-698-1 (1-1861) x ABL111881 (1-6003)

QY 1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuLe 20  
 DB 418 ATGAGCTAGTGTGAGCGCCGCTCTTCCAGGTGGCTGCAAGGAGCGCTGCAGCTAATA 477  
 QY 21 AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro 40  
 DB 478 GACAAACCCCACTTCCGAAAGAGGTGATGATCATACTCAAGTCCAGAGCAACCCAGCG 537  
 QY 41 ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60  
 DB 538 GTCAGAACCCGCGCAATTTCTTACTGTGCGCAGAGCCCTGCAGCTGAAATCGCCGACA 597  
 QY 61 GlyAlaGlyLysThrMetLysSerValValSerAlaIleValGlnGlnLysLysArgMet 80  
 DB 598 GGAGCTGCCAAGACAATGAAAGCGTGTATCCGCTGTGTGCAGCAAAAGAGCGCATG 657  
 QY 81 SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro 100  
 DB 658 TCTGAGCAGCAGCGCCCTCCCTCCAGCAGACATGCGGAGTACTGCTCTTCGCGTCCC 717  
 QY 101 AlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLysThr 120  
 DB 718 GCTGATCGGCACATCCACCTCCACAGGCTCTCTTGTGAGAAATGTATACAGACT 777  
 QY 121 ProGlnGluGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer 140  
 DB 778 CCACAAGAGAGCCCGGTATACATATACACACAGCCCTCGCAGTCTTTAAGGAAATCTAAGC 837  
 QY 141 ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuAspPheThrProLeuThr 160  
 DB 838 CCTATGAGCCGAGGAAACCTACTGACGAGTGAATGCAATCTCGGATTCACACCTCTCAC 897  
 QY 161 GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr 180  
 DB 898 GAAACCCGTGGCAAGGACAAGTACCATTTTCCGGACAAATCTGGCAGCCTGGCCACA 957  
 QY 181 ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThr 200  
 DB 958 CCACACATTAAAGGGAATGTAATATCATGTGTCTAATGATATATCGCGCGCTCGAATCACT 1017  
 QY 201 ProAspAspLeuGluAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu 220  
 DB 1018 CCAGATGATCTAGAGATCAGCCTGCCCAAAACAAACGTTCTCGATGTAAGCATTCGAG 1077  
 QY 221 ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr 240

DB 1078 ACTATCATATTTTGGTGGACACCTTGGACTGCTCCAGATCGATGGACACCGCATAG 1137  
 QY 241 ProLeuAsnLysThrThrIleValHisAlaThrHisIleThrArgAlaLeuAlaCysIle 260  
 DB 1138 CCCCTAAATAAGACAACAACCATTTGTGCATGCGCACACACAGAGCTCTGGCCCTGTATT 1197  
 QY 261 HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280  
 DB 1198 CATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1257  
 QY 281 LysArgAspIleLysLeuValGlySerProLeuArgLysLysSerGluSerMetLysAsp 300  
 DB 1258 AAGAGGACATTAAGTTGGTGGTTCACCTTACGAAAGTATTCGAGTCCATGAAGAT 1317  
 QY 301 LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsn 320  
 DB 1318 TTGTCACTTCTATCGCCACAAACTAAGTATGCAATTCAGGGTCTATGCCTAATCTAAT 1377  
 QY 321 GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlnGlnIleGlnIle 340  
 DB 1378 GAATGAAATATCCGCTCGATCGAACAGAAATAGATATACCAGAGCAGCAGATCCAGATA 1437  
 QY 341 LysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu 360  
 DB 1438 AAGGCCAAAGACTTGAATAGCTCCTCTAGTAGCAGGCTAGTTTGGCCGCCAGCAGAG 1497  
 QY 361 PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluVal 380  
 DB 1498 TTTCTATTCAACCAACAGTGCAGATCCTCGCTCAGTCCAGTCGCTTTTAAATCTCCATGA 1557  
 QY 381 GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu 400  
 DB 1558 GGTCCGAAGTCCGTGAGGAGGAGTCCGCTGAGGAATCTCTCAAGCGCCGCTCTCATGAG 1617  
 QY 401 LeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThrValAlaIle 420  
 DB 1618 TTGAGTTTTTCGGATGCACCTAGCAACGAATCATTTGACCGCAATGAAATGTAGCCATT 1677  
 QY 421 SerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla 440  
 DB 1678 TCCCTCTCTAAAAAGCAACGGGTGAGGACACTACTCTGCCCCAGAAAGTGCAGCGCGCA 1737  
 QY 441 AsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPhe 460  
 DB 1738 AATGATCTGCAAGAGCAGTAGTCCCGCCAGCTGCGCACAGCCCAATCCAAAGAGTTT 1797  
 QY 461 LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp 480  
 DB 1798 AAGCTAGCACAAACCATGTCTGATGAAGAGCCGCCACACACCAAGAGTCAAGGAGC 1857  
 QY 481 ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle 500  
 DB 1858 ACTAGCATTCAGCCTTCCGTCAGCTCTATGACTCGGAGCTGTATATGACAGCGTGCATC 1917  
 QY 501 AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr 520  
 DB 1918 AACCCGATTCATTTGCGACCACTACGCAATTTGATTCCTTTCTGGCATCTACCAATGAT 1977  
 QY 521 LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeu 540  
 DB 1978 TTGGATGAACAGGCTGTGATCGTCACTCAAGCTGACTTTTAAAGTGGTAAATGCCCTT 2037  
 QY 541 ValSerIleProAlaAspLeuAspAlaAspLeuAsnLysIleAspValGlyLysLeu 560  
 DB 2038 GTCTCCATACCCGCTGACCTGGAGCAGATTTAAATAACAAATAGAGCTTTGGTAAGCTG 2097  
 QY 561 PheAsnGluValArgAsnLysGluLeuValAlaProThrLysGluGlnSerMet 580  
 DB 2098 TTTAACAGGTCGCGCAACAAAGAGCTCGTGTGGTCCCAACAGAGGAGGAGCGTCTATG 2157  
 QY 581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600



Db 2158 AACTACCTAACGAAATACCGCTGGAGACGCTTCGTAAAGCGCGCTGTGGAGCTCTTCTTC 2217  
QY 601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
Db 2218 AGTGAGCAGATGCGCTTCGCTCATCTCCAAAGTGGCGGTATATGTCAACAAGCAAGCTCTG 2277  
QY 621 ArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
Db 2278 CGCATCCGTAGCGATCGTAATCTTCACCTAGACGTAGTTATGCAACGCACCATACTTGAG 2337  
QY 641 LeuLeuLeuCysPheAsnProLeuTyrPheLeuArgLeuGlyLeuGluValValPheGlyGlu 660  
Db 2338 CTGCTGCTTTGCTTCATCCCTTTGGCTCGCGCTTGGACTGGAGTGGTCTTTGGCGAG 2397  
QY 661 LysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
Db 2398 AAGATCCAGATGCGATCTAATCAGACATTTGAGGCGCTCAGCACCTTTATCCCAATPCG 2457  
QY 681 LeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
Db 2458 TTCTTCGGGAATTAAGTGTGAGGAGCAGAGGTACAGCAAGCATACACACTCACCGAAGAG 2517  
QY 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuProPheLeu 720  
Db 2518 TAGCGGAGACCAATTAAGAAGCACTCATTCGAGAAATCCCTCTTTTGTGCTTTTCCCTC 2577  
QY 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
Db 2578 GATCAAGCTNAGCAGAAGCGATCGTCAACACATCCCTGTTGTTGTTTAAAGTCG 2637  
QY 741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIle 760  
Db 2638 CCACATAAAGAGACCAAGGATATTCGTCTCGCTTCGTGCGAGCTGCTCGCAACATT 2697  
QY 761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
Db 2698 GGTGATATTACGGGGAACTTCGTGCGCTGGGTACGTTCTACAGCACCGCCAAACATT 2757  
QY 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
Db 2758 TTGGACAGATTTCATATATGCTTCAACAACCTTGCTGTGACCTTAGAGATGGCGTGAGA 2817  
QY 801 LeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
Db 2818 CTAACCCGAGTCATCGAGGTAAATTTTACTACGGCATCTAACCCTCGCATTTAAGGGTG 2877  
QY 821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
Db 2878 CCAGCCATCTCTCGCCTTCAGCGGATCTTCAATGTAAAGCTAGCTCTGGCGCAGCTTGGT 2937  
QY 841 GluAlaAsnPheGlnLeuGlyGlyValAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
Db 2938 GAAGCCAACTTCCAGCTAGCGCGGACATCGCGCCCAAGACATCGTTGACGACATCGT 2997  
QY 861 GluLysThrLeuSerLeuLeuTyrPheGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
Db 2998 GAGAAGACGCTTTCCTCTGCTCTGGCAACTTATTTACAAATTCGCTCGCCCAAGTTTCAT 3057  
QY 881 AlaAlaAlaThrValLeuGlnLysTyrTrpArgArgHisTrpLeuHisValValIleGln 900  
Db 3058 GCGCGCGGCACCGGTGTCCAGAAATGGTGGCGCGCTACCTGGCTGCACGTTGTTATCCAG 3117  
QY 901 ArgArgIleArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
Db 3118 CGTCGATTCGCCACAAAGAGCTTATGGCTCGCCACCGGGCCGCTACTGTTCATTCAGGCC 3177  
QY 921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
Db 3178 GTGTTCCGTGGCCACCAAGATGAGAAAGTACGTGAAGTTGTTTAAAGCGGAACGCACTCAG 3237  
QY 941 AlaAlaIleIleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960  
Db 3238 GCCGCATATATCTGCAAAAAGTTTCACCCGTCGGTATTTTGCCCCAGAGACGCTATATCAG 3297

QY 961 SerTyrHisSerIleIleThrIleGlnArgTyrTrpArgAlaGlnGlnLeuGlyArgGln 980  
Db 3298 AGCTATACACAGTATATATCATCCATCCAGCGCTGGTGGCGAGCCCAACAACACTGGAGGCAG 3357  
QY 981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArg 1000  
Db 3358 CACCGCCAGCGGTTTGTGGAGCTCCGAGAGGCTCGCATCTTTCTTCAGCGGATCTGGCGG 3417  
QY 1001 ArgArgLeuPheAlaLysLysLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
Db 3418 CGACGACTCTTTGCCAAAAAATAATTTGGCGCGCGGAAACACGCCAGACTTCAGCGCATCG 3477  
QY 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGly 1040  
Db 3478 CAAAAACAACAGCAGCTGCTAGTATATTCAAATGCATGGCGAAGCTATCATCTGGCG 3537  
QY 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
Db 3538 AGAATTCAGCGACAGCAGCTTCCTGCGCCAGAGGAGCCTCATCATCTGTTGTTTCAGCGCAGG 3597  
QY 1061 MetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
Db 3598 ATCGGAAGCAAGTGGAGTATGCTGGAGCAGCGCAGAGTTCAGCAACTAAAGCGTGCA 3657  
QY 1081 AlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
Db 3658 GCTATAAATATCCAAACAACGCTGCGCGAGCAAGCTTTCANTGAGAAGTGCACGCTGAT 3717  
QY 1101 TyrLeuAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
Db 3718 TATTTGGCACTTCGTTCCAGCGTTCCTTAAAGTTTCAGGCTTTACAGGAAAGCACAAATCCAG 3777  
QY 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
Db 3778 ATGAGAATAGATCGTAAATCACTACTATTCCTCGGAAAAAATGTATCTGCTCGTGCACAG 3837  
QY 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
Db 3838 CGACTCAGGGCCATCATGAAATGCGGAAACAGAGGGGAAAAATTTATCTGAGGCTCGCAAT 3897  
QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnMetIleGlnAspArgAsn 1180  
Db 3898 GCTTCAATACTAGTTCAAAAACGCTACCGCATCGTCAACAATAATGATCCAGGATAGAAAT 3957  
QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
Db 3958 GCATATTTAAGAACCCGCAATGTATCAATGTCCAGAGGCGCTGGAGGCCACTCTG 4017  
QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
Db 4018 CAGATCGCTCGAGAAAGGAAGAACTACCTTCATCTCAAAACGACAAACCAAAATTCAA 4077  
QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
Db 4078 ATCAAGTTCGTCGCCAAGCGCAATGAAAAAGCAAAAGACCGGAGTTTCTTCAGCTGAAA 4137  
QY 1241 LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
Db 4138 AAGGTTTACCTTGTAGTCCAGAAACGTTGGCGGGCTTTGCTGCAGATGCGAAGAGGAGCGC 4197  
QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
Db 4198 CAGGAGTACCTACATCTGCTGCGGAGGTTACCATAAAGCTGCAGCGCAGATTCATGTCTCAA 4257  
QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
Db 4258 AAATCAATGCGTTCATGCGAGCAAGTACCCGCGCACCCAGGCTGCTGTGAGCTGCTG 4317  
QY 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320  
Db 4318 CAGATGCATTTGGCGTAAATCATCTCTTTAGGAAACGGGAGGAGAACAGTTTCTTACAACCTG 4377



1321	QY	ArgGlnAlaAlaIleThrLeuGlnArgA	gTyrArgAlaAra	gLeuAsnMetIleIysGln	1340
4378	Db	CGTCAAGCAGCAATAACA	CTGCAGCGACGCTACCGAGCTCGTCTGAATGATGATCAAGCAG	4437	
1341	QY	LeuIysSerTyrAlaGlnLeuIysGlnAlaIleThrIleGlnThrArgTyrArgAla	1360		
4438	Db	TTGAAGAGTTAGCCCGAGCTGAACAGGAGCGCTATTACCATTCAAACCCGATATAGAGCC	4497		
1361	QY	LysLysAlaMetGlnIysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLys	1380		
4498	Db	AGAAGCGCAATGCAAAGACAGGTGGTCTTGTCACAAAGCAAGGAAGCCATTATCAAA	4557		
1381	QY	ValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnIys	1400		
4558	Db	GTGCAACGACGATACCGCGCAATCTGGAGATGAGGAAGCAGATTCAAGTCTACCAAAA	4617		
1401	QY	GlnArgGlnAlaValIleArgLeuGlnIysTyrTyrArgSerIleArgAspMetArgLeu	1420		
4618	Db	CAGCGCCAGCAGCTCATCGCTTGCGAAATGGTGGCGAGTATACGCACATGCGGCTG	4677		
1421	QY	CysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArg	1440		
4678	Db	TGCAAAAGCGGCTACCGAAGGATTCGACTCAGTTCAITGAGCAATTCACGCAAGTGGCGG	4737		
1441	QY	AlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArg	1460		
4738	Db	GCCACAGTCAGCGCTCGTCGCCAACGGGAGATCTTCTTAAGCACCATTCGCCAAGTGCGA	4797		
1461	QY	LeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnIleArgGluPheGlu	1480		
4798	Db	CTTATGCAGGCAATTTATCAGAGCAACTTACTGATCGTCAGCAGCGCAGGGAAITTCGAG	4857		
1481	QY	MetIysArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeu	1500		
4858	Db	ATGAGCGAAGGCTGCCGTAGTGATCCAGCGCGGTTCGTGCTCGCTGCTATGCTA	4917		
1501	QY	LysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPhe	1520		
4918	Db	AAGCGCAGACAGGATTACCAATTAATCCATCTCTGTGATCTCGTGCAGCGCAATTC	4977		
1521	QY	ArgAlaAsnArgSerMetIysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAla	1540		
4978	Db	CGTGTCAATCGCAGCATGAAGCAGGACGCGCAGGAATTTGTCCAGCTGGCTACTATTGCA	5037		
1541	QY	ValHisLeuGlnGlnIysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPhe	1560		
5038	Db	GTCCATTGGCAACAAGATTCCGTGGCAAGCTCTTAATGATGAGCAGCGTAAITGTTTC	5097		
1561	QY	GlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArg	1580		
5098	Db	CAACTGTCCGTGTTCCATGCCCGGCTTCAGGCACGTGCCCGCGCTTATGGCTCGC	5157		
1581	QY	LysArgPheGlnAlaLeuMetThrProGluMetMetAspIleuIleArgGlnIysArgAla	1600		
5158	Db	AAACGATTCAGGCCCTTGATGACACCCGAGATGATGGACCTCATTCGCCAGAGCGGCC	5217		
1601	QY	AlaLysValIleGlnArgTyrTyrArgGlyTyrLeuIleArgArgArgGlnIysHisGln	1620		
5218	Db	GCCAAGTTTATACAGCGTTACTGGCGGGCTACTTAATCCGACGACGTGAGAAGCACCAG	5277		
1621	QY	GlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsn	1640		
5278	Db	GGTCTCTTGGATATCCGGAAGCGTATTGCTCAGCTGGCGCAAGAGGCAAGGCGGTAAAC	5337		
1641	QY	SerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSer	1660		
5338	Db	TCTGTGCGCTGCAAAGTCCAGAGGCGGTGCTTCTTCCGCGAGCGCTTTATTCGATCA	5397		
1661	QY	AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu	1680		
5398	Db	GATCGGTGACAGTCTAAGTCGATGGATCGTCTTTCCGCGCATCTGTCACACCTGTCTC	5457		
1681	QY	MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg	1700		

Db	5458	ATGTGGTGTTCGGAGTTTCATGTCACAGTTTGGCTATGGCATCATGGCTCAGGCCATTGCA	5517
Qy	1701	SerGluValAspLysGlnLeuLeuLeuArgCysSerArgIleLeuLeuAsnLeuAlaArg	1720
Db	5518	TCAGAGGTGGTATAGCAGCTTATAGAGCGCTGCAGCCGGATCATCTAAATTTGGCCCGC	5577
Qy	1721	TyrAsnSerThrThrValAsnThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMet	1740
Db	5578	TACAATAGCACACCGGTGAACACGTTCCAGGAGGCGGTTTGGTCAACATTGGCCAGATG	5637
Qy	1741	LeuLeuArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrp	1760
Db	5638	TTATTTGGCTGGTCCGCAAGACAGTGAATATCAACATTTTGTGACCCCTCATTTGG	5697
Qy	1761	ValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGlu	1780
Db	5698	GTATTCGCCCACTGCTCTAAAGCGAAGATCAATTCAGACTATATGACCAACCCAGAG	5757
Qy	1781	AlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGln	1800
Db	5758	GCCATTTACATGGTGCAGCAAACTAAGAAAGCTCGTGGCCCGCAAGAAAGATGAAGCAA	5817
Qy	1801	AsnAlaArgLysProProMetThrSerClyArgTyrLysSerGlnLysIleAsnPhe	1820
Db	5818	AATGACGCAAGCGCGCCCAATGACAAGTGGACGGTATAAGGCCAAAGATAAATTC	5877
Qy	1821	ThrProCysSerLeuProSerLeuGluProAspPheGlyIleIleArgTyrSerProTyr	1840
Db	5878	ACGCGGTGTTCCCTGCCAGCTGGAGCCGGACTTCGGAAATCATCCGTACAGTCCCTAC	5937
Qy	1841	ThrPheIleSerSerValTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMet	1860
Db	5938	ACGTTTATCTCGTCCGTTACGCCCTTCGATACGATTTTGTGCAAGCTGCAGATGCATG	5997
Qy	1861	Phe 1861	
Db	5998	TTT 6000	
RESULT 3			
ABL11880			
ID	ABL11880 standard; cDNA; 8330 BP.		
XX			
AC	ABL11880;		
XX			
DT	26-MAR-2002 (first entry)		
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 30122.		
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEXE ) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB6777.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signaling and cell-cell		
PT	interactions.		

XX

PS Claim 1; SEQ ID NO 30122; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins (AB057737-  
 CC AB072072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 8330 BP; 2321 A; 2009 C; 1949 G; 2051 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 8330  
 Score: 9386.50 Matches: 1855  
 Percent Similarity: 96.32% Conservative: 4  
 Best Local Similarity: 96.11% Mismatches: 2  
 Query Match: 98.66% Indels: 69  
 DB: 4 Gaps: 3

US-09-914-698-1 (1-1861) x AB111880 (1-8330)

QY 1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeu 20  
 DB 1542 ATGGAGCTAGTGTGGAGCCCGCTCTTGTAGTGGCTGCGAGAGAGCGCTGCGAGTAATA 1601  
 QY 21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40  
 DB 1602 GACACCGCAACTCCGAAAGAGGTGATGATCATCTCAAGTCCAAAGACCAACGCG 1661  
 QY 41 Val----- 41  
 DB 1662 GTCAAGGTGAGTCAAAATGCAATTTGCCACCCCTTTTACTTACATACCCCTTTTGCAT 1721  
 QY 42 -----LysAsnProArgLysPheProThrValGlyLysThrLe 54  
 DB 1722 GTATCTCTTACCTTCTACTGAGAACCGCGCAATTTCTTACTGTGTGGCAAGACCCCT 1781  
 QY 54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValIsetralalaVa 74  
 DB 1782 GCAGCTGAAATCCCGACAGAGCTGGCAAGACAATGAAAGCGTGGTATCCGCTGCTGT 1841  
 QY 74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94  
 DB 1842 GCAGCAAAAGAACGCGATGCTGCAGCAGCAGCGCTCCCTCCAAAGCAGACATGGCGAGT 1901  
 QY 94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValGI 114  
 DB 1902 GACTGTCCTTCGCTCCGCTGCATGGGACATCCACCTCCACAGGCTCTCTTGTGCA 1961  
 QY 114 uLysAsnValTyLysThrProGlnGluProValTyIleSerProGlnProArgSe 134  
 DB 1962 GAAGAATGTATACAGACTCCCAAGAGAGCCGTTATACATATCACCACAGCCTCGCAG 2021  
 QY 134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154  
 DB 2022 TCTTAAGGAAATCTAAGCCCTTATGACGCGAGAAACCTTACTCGAGTGAATGACAAATCT 2081  
 QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174  
 DB 2082 GCGATTCACACCTCTCACCGAAACCGTGGCAAGGACAGCTACCATTTTCCCGACAA 2141  
 QY 174 nLeuAlaAlaTrpProThrProThrLysGlyAsnValLysSerCysAlaAsnAspMe 194  
 DB 2142 TCTGGCAGCGCTGGCCACACCAACACTTAAAGGGAATGTAAATCATGTGCTTAATGATAT 2201  
 QY 194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214  
 DB 2202 GCGGCCGCGTGCATCACTCCAGATGATCTAGAAGATCAGCTTGCACAAACAAACGTT 2261

QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArgI 234  
 DB 2262 CGATGTAAAGCATTCCGAGACTATCAATATTTCTGGACACCTTGGACTGCTCCAGAT 2321  
 QY 234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisTh 254  
 DB 2322 CGATGACAAACCGCATAGCCCTTAATAAGACAAACCATTTGTCATGCCACGCACAC 2381  
 QY 254 rArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThrProThrLy 274  
 DB 2382 CAGAGCTCTGGCCTGTATTTCATGAGGAGGAGGACCAAGTCCACCTAGGACGCCACGAA 2441  
 QY 274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294  
 DB 2442 GAGCGCATACACACCTGAGAGGAGACATTAAATTGGTGGGTTACCCCTTACAAAGTA 2501  
 QY 294 rSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyAlaIleGlnI 314  
 DB 2502 TTCGAGTCCATGAAAGATTTGTCACTTCTATCGCCACAAACTAAGTATGCCATTCAAGG 2561  
 QY 314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyTrpGI 334  
 DB 2562 GTCTATGCCCTAATCTAAATGAAATGAAATCCGCTCGATCGAAACAGATATATACCA 2621  
 QY 334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSe 354  
 DB 2622 GGAGCAGCAGATCCAGATTAAGCCAAAGACTTGAATAGCTCTCTAGTAGCGAGCTAG 2681  
 QY 354 rLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374  
 DB 2682 TTTGGCGGCCAGCAGAGTTTCTATTCAACACACAGTGAGATCCTCGCTCAGTCCAGTCG 2741  
 QY 374 gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi 394  
 DB 2742 TTTTAATCTCCATGAAGTAGTGGGAAGTCGGTGAAGGGAAGTCCGCTGAAGAATCTCTCA 2801  
 QY 394 sLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGlnSerLeuTrAr 414  
 DB 2802 CAAAGCCCGCTCTCATGAGTTGAGTTTTTCGGATGCACCTAGCAAGATCAATGTACCG 2861  
 QY 414 gAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr 434  
 DB 2862 CAATGAAACTGTAGCCATTTCCTCTTAAAGCAACGGTTGAGGACACACTACTCTGCC 2921  
 QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerSerAlaHiAlaTrpProHi 454  
 DB 2922 CAGAAGTCAGCGCCGCGCAATGCAATCTCAAGAAGCACTAGTGCACCGCTGGCCACA 2981  
 QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
 DB 2982 CGCCCAATCCAGAGTTTAAGCTAGCAACCACTGTCTAGTAGAAGAGAGGCCGCCAC 3041  
 QY 474 rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyAspSerGluLe 494  
 DB 3042 ACCACGAAAGTCAGGACACTAGCATTTCAGCTTCGCTCCGCTCTATGACTCGGAGCT 3101  
 QY 494 uTyThrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrIleAspProPh 514  
 DB 3102 GTATATGACAGCTGCATCAACCCGATCCATTTTCAGCAACTAGCACAATATGATCCATT 3161  
 QY 514 eLeuAlaSerThrMetTyLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy 534  
 DB 3162 TCTGGCATCTACATGATTTTGGATGAACAGGCTGTGGATCGTTCATCAAGCTGACTTTAA 3221  
 QY 534 sLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsnLy 554  
 DB 3222 AAAGTGGTTAAATGCCCTTGTCTCCATACCCGCTGACCTGGAGCGCAGATTTAAATAACAA 3281  
 QY 554 sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValValAlaProTh 574  
 DB 3282 AATAGACGTTGTAGCTGTTTAAAGAGTGCACAAAGAGCTCGTGTGTGCTCCCCAC 3341

Qy 574 rLysGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAl 594  
Db 3342 CAAGGAGAGCGTCTATGAACCTACCTACGAAATACCGCTGGAGAGCGCTTCGTAAGGC 3401  
Qy 594 aAlaValGluLeuPheSerGluGlnMetArgLeuProCysSerTyrValAlaValTy 614  
Db 3402 GGCTGTGGAGCTCTTCTCACTGAGCAGATCGCTCCAAAGTGGCGGTATA 3461  
Qy 614 rValAsnLysGlnAlaLeuArgLysLeuArgSerAspArgAsnLeuHisLeuAspValValMe 634  
Db 3462 TGTCAACAAGCAAGCTCTGCGCATCCGTAAGATCGTAATCTTCACTAGACGTAGTTAT 3521  
Qy 634 tGlnArgThrTyrLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLysLysLe 654  
Db 3522 GCAACGACCACTTGGAGCTGCTTGGCTTCAATCCCTTTGGCTGGCGCTTGGACT 3581  
Qy 654 uGluValValPheGlyGluLysLysLeuGlnMetGlnSerAsnArgAspLysValGlyLeuSe 674  
Db 3582 GGAAGTGGTCTTTGGCGAGAAAGATCCAGATCGAGTCTAATCGAGACATTTGTAGGCTCAG 3641  
Qy 674 rThrPheLeuAsnArgLeuPheArgAsnLysCysGluGlnArgTyrSerLysAl 694  
Db 3642 CACCTTTATCTCAATCGCTTGTTCGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC 3701  
Qy 694 aTyrThrLeuThrGluGluTyrAlaGluThrLeuLysLysHisSerLeuGlnLysLysLe 714  
Db 3702 ATACACACTACCGAAGAGTAGCGGAGACCATTAAGAACGACCTCATTTGCAGAAATCCT 3761  
Qy 714 uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgLysValLysHisAsnProCy 734  
Db 3762 CTTTTTGTCTTTTCTCGATCAAGCTAAGCAGAGCGCATCGTCAAGCAACATCCCTG 3821  
Qy 734 sLeuPheValLysLysSerProHisLysGluThrLysAspLysLeuLeuArgPheSerSe 754  
Db 3822 TTTCTTTGTTAAAGTGCACACATAAGAGACCAAGGATATTCCTGTCGCTTCTCGTC 3881  
Qy 754 rGluLeuLeuAlaAsnLysLysLeuArgLysLysLeuArgLysLysLysLysLysLys 774  
Db 3882 GGAGCTGCTGCCAACATTTGTTGATTAATACCGGGAATCTGTCGCTGGGCTACGTTCT 3941  
Qy 774 uGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValas 794  
Db 3942 ACAGCAGCGCAACATTTTGGACGAGTTCGATATTCCTTCAACACTTGGCTGTGA 4001  
Qy 794 pLeuArgAspGlyValArgLeuThrArgValValGluValLysLeuLeuArgAspLys 814  
Db 4002 CCTAAGAGATGCGGTGAGCTAAACCGAGTATGAGGTAATTTTACTACGCGATGATCT 4061  
Qy 814 uThrArgGlnLeuArgValProAlaLysSerArgLeuGlnArgLysPheAsnValLysLe 834  
Db 4062 AACCGCCAGTTAAGGTGCCAGCATCTCTCGCTTCAGCGGATCTTCAATGTAAAGCT 4121  
Qy 834 uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspLysLysAlaAlaGlnAs 854  
Db 4122 AGCTCTGGCGCACTTGTGTGAAGCAACTTCCAGCTAGCGGCGACATCGCGCCCAAGA 4181  
Qy 854 pIleValAspGlyHisArgGluLysThrLeuSerLeuLeuThrGlnLeuLysLysPhe 874  
Db 4182 CATCGTACGCGACATCGTGTGAGAGACGCTTTCCTGCTCTGGCAACTTATTTACAAAT 4241  
Qy 874 eArgSerProLysPheHisAlaAlaAlaThrValLeuGlnLysTyrTrpArgArgHisTr 894  
Db 4242 CCGCTCGCCCAAGTTTCAATGCGCGGCCACGCTGCTCCAGAAATGGTGGCGCGTCACTG 4301  
Qy 894 pLeuHisValValLysGlnArgArgLysGlnLysGlnLeuMetArgArgHisArgAl 914  
Db 4302 GCTGACGCTGTATCCAGCGTCCGCAATTCGCCACAAAGAGCTTATGCTGCCACCGGCG 4361  
Qy 914 aAlaThrValLysGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPh 934  
Db 4362 CGCTACTGCTCAATTCAGGCGCTGTTCCGTGGCCACCGACAGATGAGAAAGTACGTGAAGTTGT 4421  
Qy 934 eLysThrGluArgGlnAlaAlaLysLysPheThrArgArgTyrLeuAl 954

Db 4422 TAAGCGGAAGCACTCAGCGCGCAATAATCTCCAAAGTTCCCGCTCGTATTGGC 4481  
Qy 954 aGlnLysGlnLeuTyrGlnSerTyrHisSerLysLysLeuGlnArgTyrTrpArgAl 974  
Db 4482 CCAGAGCAGCTATATCAGAGCTATCAAGTATATCACCATCCAGCGCTGGTGGCAGC 4541  
Qy 974 aGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaLysPhe 994  
Db 4542 CCAACAATCGGAAGGAGCACCAGCGGCTTGTGGAGTCCGAGAGGCTGGATCTT 4601  
Qy 994 eLeuGlnArgLysLeuPheArgLeuPheAlaLysLysLeuLeuAlaAlaLysLys 1014  
Db 4602 TCTTCAGCGGATCTGGCGGCGACCTCTTTCGCAAAAACCTATTTGGCGCGCGGAAC 4661  
Qy 1014 rAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaLysSerTyrLysGlnMetGlnTr 1034  
Db 4662 AGCCAGACTTCAGCGATCGCAAAAACCAACAGCGAGCTCTAGTTATATTCAAATGCAATG 4721  
Qy 1034 pArgThrTyrGlnLeuGlyArgLysGlnArgHisGluPheLeuArgGlnArgAspLysL 1054  
Db 4722 GCGAAGCTATCAGCTGGCGAGATTCAGCGACAGCGAGTTCTTGGCGAGAGGACCTCAT 4781  
Qy 1054 eMetPheValGlnArgArgMetArgSerLysTyrSerMetLeuGluGlnArgLysGluPh 1074  
Db 4782 CATCTTTGTTTCAGCGCAGGATCGAGCAAGTGGAGTATGCTGGAGCAGCGCAGGAGTT 4841  
Qy 1074 eGlnGlnLeuLysArgAlaAlaLysLysGlnGlnArgTrpArgAlaLysLeuSerMe 1094  
Db 4842 CCAGCAACTAAAGCGTGCAGCTATAAATATCCAAACACGCTGGCGAGCGAAGCTTTCAT 4901  
Qy 1094 tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTy 1114  
Db 4902 GAGAAAGTGCACGCTGATTTATTTGGCACTTCGTTTCAGCGCTTCTTAAAGTTTCAGGCTTA 4961  
Qy 1114 rArgLysAlaThrLysGlnMetArgLysLysLysLysLysLysLysLysLysLysLys 1134  
Db 4962 CAGGAAGCCACATCCAGATGAGATAGATCGTAATCACTACTTATTCCTTCGCGAAGAAA 5021  
Qy 1134 nValLysCysLeuGlnArgLeuArgAlaLysMetLysMetArgGluGlnArgGluAs 1154  
Db 5022 TGTTATCTGCTGCAACAGCGACTGAGCGCCATCATGAAATGCGCGAAGCAAGAGGAAA 5081  
Qy 1154 nTyrLeuArgLeuArgAsnAlaSerLysLysLysLysLysLysLysLysLysLysLys 1174  
Db 5082 TTATCTGAGGCTGGAATGCTTCAATAGTTAGTTTCAAAACGCTACCGCATCGCTCAACA 5141  
Qy 1174 nMetLysGlnAspArgAsnAlaTyrLeuArgThrArgLysCysLysLysLysLysLys 1194  
Db 5142 AATGATCCAGATAGAAATGATATTTAAGAACCGCAATGTATCATCAATGTCAGAG 5201  
Qy 1194 gArgTrpArgAlaThrLeuGlnMetArgArgLysLysLysLysLysLysLysLysLys 1214  
Db 5202 GCGCTGGAGAGCCACTCTGAGATCGCTGCGAAGGAAAGAACTACCTTCATCTCCAAAC 5261  
Qy 1214 rThrThrLysArgLysGlnLysPheArgAlaLysArgGluMetLysLysLysLysLys 1234  
Db 5262 GACAAACCAACGAAATCAAATCAAGTTCCGTCGCAAGCGCAAGTAAAGAAAGCAAGAGC 5321  
Qy 1234 aGluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLe 1254  
Db 5322 CGAGTTCTTCAGCTGAAAAGGTTACCTTGTAGTCCAGAAACGTCGCGCGGCTTGTCT 5381  
Qy 1254 uGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrLysLysLeuG 1274  
Db 5382 GCAGATCGAAAGGAGCGCCAGGAGTACCTACATCTGCGGAGGTACCATAAAGCTGCA 5441  
Qy 1274 nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrG 1294  
Db 5442 GCGAGATTCATCTCAATCAAAATCAATCGGCTTCTGCGAGCCCAAGTACCGCGCGCCCA 5501  
Qy 1294 nAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluAr 1314

Db 5502 GGCTGCTGTGAGCTGCTGCAGATGCATTGGCGCTAATCATCTCTTAGGAAACGGGAGAG 5561  
QY 1314 gAenSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaAr 1334  
Db 5562 GAACAGTTTCTTCAACTGGTCAAGCAGCAATAACACTGCAGCGAGCGTACCAGCTCG 5621  
QY 1334 gLeuAenMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIl 1354  
Db 5622 TCTGAATATGATCAACAGCTTGAAGAGTTACGCCAGCTGAACAGCAGCTATTACCAT 5681  
QY 1354 eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGl 1374  
Db 5682 TCAACCCCGATATAGAGCCCAAGAGGCAATGCAAAAGCAGAGTGTCTTGTACCAAAAGCA 5741  
QY 1374 nArgGluAlaIleLysValGlnArgArgTyrArgGlyAsnLeuGlnLeuMetArgLysGl 1394  
Db 5742 AAGGGAAGCCATTATCAAGTGCACACGACATACCGCGCAATCTCGAGATGAGGAAGCA 5801  
QY 1394 nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrPrpArgSe 1414  
Db 5802 GATTGAAGTCTACCAAAACAGCGCCAGGCGAGTCATCCGCTTCAGAAATGGTGGCGCAG 5861  
QY 1414 rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe 1434  
Db 5862 TATACCGGCATCGCGCTGTGCAAGCGGGCTACCGAAGGATTTCGACTCAGTTTCATTGAG 5921  
QY 1434 rIleGlnArgLysTyrArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe 1454  
Db 5922 CATTCACGCAAGTGGCGGCCACAGTCGAGGCTGTGCGCAACGGGAGATCTTCTTAAG 5981  
QY 1454 rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl 1474  
Db 5982 CACCATCCGCAAAAGTCGACTTATGAGCGCATTTATCAGAGCAACTTTACTGATGCGTCA 6041  
QY 1474 nGlnArgArgGluPheGlnMetLysArgArgAlaAlaValValIleGlnArgArgPheAr 1494  
Db 6042 GCAGCGCAGGGAATTCAGATGAAGCAAGGCTGCGGTAGTATCCAGCGCGGTTTCG 6101  
QY 1494 gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl 1514  
Db 6102 TGCTCGCTGTGCTATGCTTAAGCGGAGACAGGATTACCAATTATTCATTCCTCTGTGAT 6161  
QY 1514 eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa 1534  
Db 6162 CCTGTGTGACGCGCAATTCGTGCTAATCGCAGCATGAAGCAGCGCACCGCAAAATTGT 6221  
QY 1534 lGlnLeuArgThrIleAlaValHisLeuGlnLysPheArgGlyLysArgLeuMetIl 1554  
Db 6222 CCAGCTGCGTACTATTGCGATGCTCATTTGCAACAAAAGTTCCGFTGGCAAGCGTCTAATGAT 6281  
QY 1554 eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl 1574  
Db 6282 TGAGCAGCGTAATTGTTTCCAACGTCTCGCTGTTCATGCGCGGTTTCAGGCGACGTGC 6341  
QY 1574 aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe 1594  
Db 6342 CCGCGGCTTATGCTTCGCAAAAGATTCCAGGCGCTGATGACACCGAGATGATGACCT 6401  
QY 1594 urIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrIleuIleAr 1614  
Db 6402 CATCCGCCAAGACGCGGCCCAAGGTTATACAGCGTTACTGCGGGGCTATCTAATCCG 6461  
QY 1614 gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGl 1634  
Db 6462 ACGACCTCAGAGCACACAGGGTCTCTTGATATCCGGAAGCGTATTGCTCAGCTGCGGCA 6521  
QY 1634 nGluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuAr 1654  
Db 6522 AGAGGCAAAAGCGCGTAAATCTCTGCGCTGCAAAAGTCCAGGAGCGGTGCTTCTCTCG 6581  
QY 1654 gGlyArgPheIleAlaSerAspAlaLeuAlValLeuSerGlnLeu- 1669  
Db 6582 CGACCGCTTATCGCATCAGATCGGTTAGCAGTGCCTAAGTTCGATTGGGTAAGTGATCCAG 6641

QY 1670 -----AspArgLeuSerAr 1674  
Db 6642 CTACGTTAGTCTCTATTATATAACTAATCCTGCTTAATAATCCCGCAGATCGTCTTTCCG 6701  
QY 1674 gThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIl 1694  
Db 6702 CACTGTGCCACACCTCTCATGTGGTGTTCGGAGTTTCATGTCACGTTTTGCTATGGCAT 6761  
QY 1694 eMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgIl 1714  
Db 6762 CATGGCTCAGGCCATTTCGATCAGAGGTGATAGCAGCTTATAGCGCTGCACCGCGAT 6821  
QY 1714 eIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGluGlyLysLe 1734  
Db 6822 CATCTTAATTTGGCCCGCTACATAGCACCCAGGTGACACACGTTCCAGGAGGCGGTTT 6881  
QY 1734 uValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluIlePheAsnTh 1754  
Db 6882 GGTCAACATTGCCAGATGTTATTGGCTGGTGCAGACAAAGACAGTGAGATATTCAACAC 6941  
QY 1754 rIleuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770  
Db 6942 TTTGTGCACCCCTCATTTGGGTATTTCGCCCACTGTCTCTAAAGCGAAAGGTAGCTAAATA 7001  
QY 1771 -----IleIleH 1773  
Db 7002 GCGAGAGGATTAGTTTTAGCGAGTTTTTAATCATATTCTTAATGCTTCCAGATCATTC 7061  
QY 1773 iAspTyrMetThrAsnProGluAlaIleTyrMetValArgGluThrLysLysLeuVala 1793  
Db 7062 ACGACTATATGACCAACCCAGAGGCCATTATACATGTTGCGGCAAACTAAGAAGCTCGTG 7121  
QY 1793 lAargLysGlnLysMetLysGlnAsnAlaArgLysProProMetThrSerGlyArgT 1813  
Db 7122 CCCGCAAGGAAAGATGAAGCAAAATGCACGCAAGCCGCGCAATGACAAAGTGCAGCT 7181  
QY 1813 rLysSerGlnLysIleAsnPheThrProCysSerLeuProSerLeuGluProAspPheG 1833  
Db 7182 ATAAGAGCCAAAGATAACTTCACGCCGTTCCTGCCAGCTTGGAGCGGACTTCG 7241  
QY 1833 lYlleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAspThrIleL 1853  
Db 7242 GAATCATCGCTACAGTCCCTACAGCTTATCTCGTCCGTTTACGCTTTCGATACGATT 7301  
QY 1853 euCysLysLeuGlnIleAspMetPhe 1861  
Db 7302 TGTGCAAGCTGCAGATCGACATGTTT 7327  
RESULT 4  
AAH14237  
ID AAH14237 standard; cDNA; 2453 BP.  
XX AAH14237;  
XX 26-JUN-2001 (first entry)  
DT Human cDNA sequence SEQ ID NO:11532.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
KW Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
PD 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.

303	Db	GCTAGCGGGTTAAATGTTTCGAAAAGATAGACACCTATGGAAGAAATGCGGAGAACGTCAG	362
637	Qy	ThrIleuLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluVal	656
363	Db	AAAGTCTCGAATTGGCTGTGTCTCAAAATCCITTTGTGGCTCGAATTGGTCTAGAGACA	422
657	Qy	ValPheGlyGlyLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe	676
423	Db	ACTATATGGGAATCTCATATCTTTTGGGAAGATAACAGTGAATGTCACAGGGTTGGCTATGTTT	482
677	Qy	IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys	693
483	Db	ATTCTGAATCGCCTACTTTGGAAATCTGTATATAGCAGCTGAGTATAGACACCCACATGTT	542
694	Qy	AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle	713
543	Db	CCTCACCTGTATAGAGATGTCATGAAGAAGCTTTGTCCAAGTTTACATTTGAAAAAGTTA	602
714	Qy	LeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro	733
603	Db	TTGTGTGTGCTGTGTTTCTTGATATATGCTPAAATTTCCAGACTCATTTGATCATGATCCCT	662
734	Qy	CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer	753
663	Db	TGTCCTCTCTGTAAAGATGCCGAATTCAGGCTAGTAAAGAAATCTTTTAGGCTTTTCA	722
754	Qy	SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrVal	773
723	Db	CGAANTTCTTAAGTGTGAAGGTGACCTTTCCGTCGCCTTGGCTATTGGGATTAACCT	782
774	Qy	LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal	793
783	Db	GTAAACCATGTTCCAGACACCATTTTGATGAATTTGCAATTTGCGGTACAAATCTTCCGCTA	842
794	Qy	AspLeuArgAspGlyValAlaLeuThrArgValValGluValIleLeuLeuArgAspAsp	813
843	Db	GACTTGAATGTGGAGTGGCCCTGTGCGAACCATCGAATCTCTCACAGAAGCTGGGAC	902
814	Qy	LeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLys	833
903	Db	CTCTCAAGAACTCAGGATTCGGCAATPAGTCGTCTTCAAAAGATGCACATGTTGAC	962
834	Qy	LeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAsp	849
963	Db	ATTGTTCTTCAAGTCTTAAATCACAGGAATTTGAATTAAGTGTAGTACGATGGAATAACA	1022
850	Qy	IleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGln	869
1023	Db	ATTCTATCTAAGGATATPTGGGATAGGCACAGAGAAAAAACTCTCAGGTTGCTTTGGAAA	1082
870	Qy	LeuIleTyrLysPheArg-----SerProLysPheHis-----	875
1083	Db	ATAGCGTTTGCTTTTCAGGTGGATATTTCCTTAACTTAGATCAATTAAGGAAGAAATT	1142
875	Qy	-----	875
1143	Db	GCCTTTCTAAACACACAAAGATATAAAGAAAAACAATATCTCTACTATCATGCAATTCT	1202
876	Qy	-----SerProLysPheHis-----	880
1203	Db	GATGATCTTTATTAATAGAAAAAAGCAAGGGATAGTGGTTCTTTTGAACAATATAGT	1262
881	Qy	-----AlaAlaAla--ThrValLeuGlnLys	888
1263	Db	GAATAACATAAGTTATTTGATGGATTGGGTAATGCTGTTTGTGCCCTCTATATAATAAAG	1322
888	Qy	strip-----	889
1323	Db	GTGGAGAAATTTTACAGTGTCTTTCTCAGACGGCGGTGTGTTATGTTACCTGATCCACCAT	1382
889	Qy	-----	889
1383	Db	TAGCATCTTCCTATGTGCCATTTAGACGGCTATATGTACAGCGGTACTACTCAAACTGTGAA	1442

```
QY      890  -----Tr 890
Db      1443  TGTACGCAAACTGGTTCAGTGGTATTAATTCATCATCTCAATCTGATGACAGTTCCTG 1502

QY      890  pArgArgHisTrpLeuHisValVal
Db      1503  GATATGTCACCTTAAAGCATTTGATGATAAATACITTCAGAGCTATACAAAGAGCTCCTA 1562

QY      899  -----IleGlnArgArgIleArgHis
Db      1563  GAAATGAAAGAAAGAAATTTT-CACATTGGTTAGTCTGTCAGTTAGAGACCTTGGTGAAT 1621

QY      905  -----
Db      1622  ACCTGCTATGATTAATCATTCAGATATGTCATAATCAATTCAGATCAAAAGGTGGTTAT 1681

QY      906  -----LysGluLeuMetMetArgHisArgAlaAl
Db      1682  TACCTATTTTCTCATTTCTTTGTGCAAGGCTTTTGGATCTTCGTAAGAAATAAGAGCTGC 1741

QY      915  aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrrValLysLeuPheLy 935
Db      1742  TCGACTCATACAAACACATCGGAGAAATATAAATACTAAACACAGATCTCAACGCCATCA 1801

QY      935  sThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyrrLeuAlaGl 955
Db      1802  G---GAGAGAGAGAAGCTGCAAGATTTATCAATTCGCTGTAATCAATTTCTAGCAAA 1858

QY      955  nLysGlnLeuTyrrGlnSerTyrrHisSerIleIleThrIleGlnArgTrpArgAlaGl 975
Db      1859  ACAAGATTGAGAAAGAGCTTAATGCAGCACTCGTCAATTCAGAAATATTGG----- 1910

QY      975  nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaIlePheLe 995
Db      1910  ----- 1910

QY      995  uGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015
Db      1911  -----CGAAGAGCTTCAGCAGAGAAATTTAATGTTAAAGGA 1954

QY      1015  aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaLysTyrrIleGlnMetGlnTrpAr 1035
Db      1955  AAAGCTGGAAGAAAGTTCAAAATAAA-----GCAGCATCACTTATTCAGCGATATTGGAG 2008

QY      1035  gThrTyrrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055
Db      2009  AGATATATCC-----ACTAGACAAAGATTTCGAAATGGAATTAATTTCAAT 2056

QY      1055  tPheValGlnArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGl 1075
Db      2057  CATCTGCTCAATCTAGGATAAGA-----ATGATAATTGCTGTTACATCTTATAA 2104

QY      1075  nGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetAr 1095
Db      2105  ACCGATCTTTTGGGCTACACATTACATTCAGAGCATTCGCGTCTTATTTAAAGAGAA 2164

QY      1095  gLysCysAsnAlaAspTyrrLeuAlaLeuArgSerSerValLeuLysValGlnAla---Ty 1114
Db      2165  ACAAGATCAACAAGATATGAAATGCTGTAATCAATCACTCTTATTAATCAATCTATGTT 2224

QY      1114  rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrrSerLeuArgLysAs 1134
Db      2225  CAGAAAA-----TGGAAGCAACGTAAA-- 2246

QY      1134  nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgGl 1153
Db      2247  -----ATGCAATCAACAAGATAAAGCTCAGTAATATTCCAAAGACCTTTTGAAGA 2296

QY      1153  uAsnTyrrLeuArgGlnHisArgGlnArgPheValGlnLysArgTy 1169
Db      2297  ATGGCATTTAAGAAACAAAGCAAGTAAGAGAAATTTCTGCTATTATCATACATCATGGTA 2356
```

```
QY      1169  rArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrrLeuArgThrArgLysCysI 1189
Db      2357  TAGAATGCATAAAGAAATTA-----CGGAAGTATATTATATTAGATCTTGCT 2404

QY      1189  eIleAsnValGlnArgArgTrpArg 1197
Db      2405  TGTATCATTCAGAAAAAGATTCCGG 2429

RESULT 5
AAH14236
ID   AAH14236 standard; cDNA; 3887 BP.
XX
AC   AAH14236;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11530.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PE   28-JUL-2000; 2000EP-00116126.
XX
PF   29-JUL-1999; 99JP-00248036.
XX
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR   WPI; 2001-318749/34.
XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 11530; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 3887 BP; 1436 A; 652 C; 760 G; 1039 T; 0 U; 0 Other;
```



Alignment Scores:			
Pred. No.:	2,02e-41	Length:	3887
Score:	707.50	Matches:	276
Percent Similarity:	40.70%	Conservative:	225
Best Local Similarity:	22.42%	Mismatches:	412
Query Match:	7.44%	Indels:	319
DB:	4	Gaps:	40
US-09-914-698-1 (1-1861) x AMH14236 (1-3887)			
Qy	868	TyrGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln	887
Db	359	TGGATGATGAAGAAAGGATCGGA-----GAGATGCACAGGGCTGTCTTCTCATCCAG	412
Qy	888	LysTrpTrpArgArgHisTrpLeuHis-----Val	897
Db	413	TCTACTTTTCAGAAATGACAGATTACATATGAGATATACAGGCTTTTGAAACAGGCCTCCGTT	472
Qy	898	ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet	909
Db	473	GTGATCCACACAGCAATACCAAGCAAAATAGAGCTGCAGAACTGCAGAGCAGCATATATCTC	532
Qy	910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929
Db	533	AGACAAAGACACTCTGCTGTGATCCTTCAGGCTGCATTTCAGGGGTATGAAACTAGAAAGA	592
Qy	930	TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThr	949
Db	593	CATTGTGAAGAGTATGCATTCC-----TCTGCAACCTTTATTTCAGAGTAGGTTT	640
Qy	950	ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969
Db	641	AGATCACTACTGTCGAGGAGAGATTCATTTCCTCCCAAAAAGCTACTATTTTGTTCAG	700
Qy	970	ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGlnLeuArg	989
Db	701	AGGAAATACGAGCCACCATTGTGTGCCAAACATAAATGTACCAATTCCTGCACCTAAGA	760
Qy	990	GluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeu---	1008
Db	761	RAAGCAGCCATTACATACAGTCATCTTACAGAGACTCATGTTAAAGAAAGAGTATCAA	820
Qy	1009	-----LeuAlaAlaAlaGlnThrAlaArgLeuGlnArgSerGlnLysGln---	1023
Db	821	GAATGCAAAAGGCTGCAGT--TCTCATTCAGGCTACTTTCAGATGCACAGCAAAAAGAAA	879
Qy	1024	-----GlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThr	1036
Db	880	TATATTACATTTTCAGACTTGGAAACATGCTTCAATTCATAATTCAGCAACATATCGAACA	939
Qy	1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe	1056
Db	940	TATAGAGCTGCAAAATGTCAAAGAGAAAATATATACAGCAATGCAATTCCTGCTGTGTT	999
Qy	1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGlu	1073
Db	1000	ATTTCAGGCTGCATATAAAGGAATGAAGCAAGCAACTTTTAAAGGAAAACACAAAGCT	1059
Qy	1074	-----PheGlnGlnLeu	1077
Db	1060	TCTATTGTAATCAAGCCACCTACAGAAATGTATAGCAGTATTGTTCTTACCAAAAGCTT	1119
Qy	1078	LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097
Db	1120	CAGTGGCTTACAAAATCATCAAGAAAAATATAGAGCAATAAATAAAGAAACAGAAA---	1176
Qy	1098	AsnAlaAspTyrLeuAlaLeuArgSerValLeuLys-----ValGlnAla	1113
Db	1177	-----GTATTTCAACAATGAATTAAGAAAGAGACTTGTGTTCAGGCA	1221
Qy	1114	---TyrArgLysAlaThrIleGlnMetArgIleAsp-----	1124
Db	1222	GGTTTCAGGACATGACATATAAATAACAGATTCAGAAACAGCAGCAGCAGCTGCCATTATT	1281

Qy	1125	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1134
Db	1282	ATTTCAGAGAGCATTTGTAAGCCCTTTAAATAAGGAGCATTAATCTCCCATTTAGACACACA	1341
Qy	1135	ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	1151
Db	1342	GTAGTTTCTATTCAAAGAGATACAGAAAACCTAACTGCAGTCGCTACCAAGCAGTTATT	1401
Qy	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
Db	1402	TGTATACAGTCTTATTACAGAGGCTTTAAAGTACAGAAAGGATATTCAAAATATGCACCGG	1461
Qy	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
Db	1462	GCTGCCACACTAATTCAGTCATCTTCATCGAATGACAGGGCCAAAGTTGAT-----	1512
Qy	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
Db	1513	---TATGAACAAGAAAGAACTGCAATTCGTGTATACAGAATTTATTAGGTTGTATGTT	1569
Qy	1201	GlnMetArgArgGluArgLysAsnTyrIleuHisLeuGlnThrThrLysArgIleGln	1220
Db	1570	AGAGTAAAAACAGAAAGAAACCTTTTTCAGCAGTTCAGAAATCTGTACGAACCTATTTCAG	1629
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys	1240
Db	1630	GCTGCTTTTAGAGGATGAAA-----GTTAGACAAAATTTGAAAATGTATCAGAGGAA	1683
Qy	1241	LysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
Db	1684	AAGATGCGAGCAATGTT-----AACCAATCTGCACCTCTGCTGTTCAGAAAGTAAA	1734
Qy	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisIleGln	1280
Db	1735	ACTCAGTATGAAGCTGTTCAAAGTGAAGGTGTATGATTCAAGAGTGTGTATTAAGCTTCT	1794
Qy	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu	1300
Db	1795	GGCTTGCTGTTTCACAGAAAGCAGAGATCATCTTCAAGTAGGCTGTCAGTAACAAT	1854
Qy	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
Db	1855	CAAAAAGCTTTTGTAGATGCTCAGAGAAAACCTGGAACACACAGAAATGTCTGCCCTA	1914
Qy	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
Db	1915	CGGATTCAGTTCTTCTTCAGATGCTGTGTATCGGAGAGATTTGTTTCAGCAGAAAAGA	1974
Qy	1323	AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLys	1342
Db	1975	GCTGCTATCATCTTTACAGCAATTTTATTAGGAG-----TGGCAACACCAAGAAA	2022
Qy	1343	SerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
Db	2023	CAGTTTTTACTATATAGAAAAGCAGCAGTGGTTTTACAAAATCACTACAGACATTTCTG	2082
Qy	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
Db	2083	TCTGCAAAACATCAAGACAGCTTATTTCAGATCAGAGCAGGTGTATTCATTATTCAA	2142
Qy	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
Db	2143	GCTGAAGTAAGGATTTTATACAGAAACGGAG-----TTTCAGGAAATTTAAA	2190
Qy	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
Db	2191	AATAGCACCAATAAAATTCAGGCTATGTGGAGGATATAGAGCCAGCAAAATTTATGCT	2250
Qy	1422	Lys-----AlaGlyTyrArg-----	1426
Db	2251	AAAGTGAAGGCTGCTCGAAGATTCAAGCCTGTTATAGATTGTTGGAGAGCACACAAAGAA	2310



QY 1426 ----- 1426  
 Db 2311 TATCTAGCTATTTAAAGCTGTGTAAATTTTCAAGTTGCTTCTATACAACTAGAG 2370  
 QY 1427 -----ArglleArgLeuSerSerleuSerlleGlnArgLysTrpArgAla 1441  
 Db 2371 AGAACACGGTTTTTGAATGTGAGAGCATCAGCAATTATCATTCAGAGAAAATGAGAGCT 2430  
 QY 1442 ThrValGlnAlaArgGlnArgGlnlePheleuSerThrleArgLysValArgLeu 1461  
 Db 2431 ATACTTCCTCAAGATAGCTCATGAACACTTCTTAATGATAAAAGA----- 2478  
 QY 1462 MetGlnAlaPheleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMet 1481  
 Db 2478 ----- 2478  
 QY 1482 LysArgArgAlaValVallleGlnArgArgPheArgAlaArgCysAlaMetLeuLys 1501  
 Db 2479 ---CATCGAGCTGCTGTGTTGATCCAGCACATTTATAGAGA-----TATAAA 2523  
 QY 1502 AlaArgGlnAspTyrGlnLeuLleGlnSerSerValleLeuValGlnArgLysPheArg 1521  
 Db 2524 GGAAGCAGGCTCTTTCTCGCAGAAATCTGCTGTGATCATACAAAATATATACGA 2583  
 QY 1522 AlaAsnArgSerMetLysGlnAlaArgGlnGlnPheValGlnLeuArgThrilleAlaVal 1541  
 Db 2584 GCCAGGAGGCTGGAAGCATGAAGATAAAATATATATTGAATTTAAAAAATCTACAGTT 2643  
 QY 1542 HisLeuGlnGlnLysPheArgGly-----LysArgLeuMetlleGlnGlnArgAsn 1558  
 Db 2644 ATCCTACAGCACTGGTGGCGTGTGGCTAGTAGACGAAAGATTTTGTAGAACAGAGAGCC 2703  
 QY 1559 CysPheGlnLeuLeuArg----- 1564  
 Db 2704 AAAATTCGACTCTTCACCTTCAGCTGCATATATATACCTGAATGCTGTAGAAAT 2763  
 QY 1564 ----- 1564  
 Db 2764 CAAAGAGCCTATAAACTTTTACCTGGCTGTGAAGATGCTAAACAGCAGGTTAATTCAGTC 2823  
 QY 1565 ---CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
 Db 2824 ATCTGTATTTCAGATGGTTTCGAGCAAGA-----TTACAAGAAAAGAGATTT 2871  
 QY 1584 ---GlnAlaLeuMetThrProGluMetMetAsp-----LeulleArg 1596  
 Db 2872 ATTCAGAAATATCATAGATCAAAAGATGAGCATGAGCTCAAGATGCTGAGCCAG 2931  
 QY 1597 GlnLysArgAlaAlaLysVallleGlnArgTyrTrpArgGlyTyrleulleArgArg 1616  
 Db 2932 CGAAATAGGGCTGCATCAGTAAATACAGAAAGCAGTCGCCATTTTCTCTCCGTAAAG 2991  
 QY 1617 GlnLysHis-----GlnGlyLeuLeuAspIleArg----- 1626  
 Db 2992 CAGGAAAAATTCATAGTGAATCATATAAAATTCAGGCATATFAGGAGGCTATTCTTGG 3051  
 QY 1627 -----LysArgilleAlaGlnLeuGlnGlnAlaLysAlaValAsn 1640  
 Db 3052 AGGAGAAAAATCATTTGATCAAAAAATTAAGCTATAGCTAAGTCTTCAAGTTGTTAAT 3111  
 QY 1641 SerValArgCysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheilleAlaSer 1660  
 Db 3112 AGG-----GAGATTCGAGAGAAAAACAACACTCTACAAAAGAACTGCACCTGCAC 3162  
 QY 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgArgLeuSerArgThrValProHisLeuLeu 1680  
 Db 3163 CATTACCTTTTGACATATAGCACCTTTCTGCCATTCCTGAGGCCCTTAAACACCTAGAG 3222  
 QY 1681 MetTrpCysSerGluPheMetSerThrPheCysTyrGlylleMetAlaGlnAlaIleArg 1700  
 Db 3223 GTAGTTACTAGA-----TTGCTCCACTTTGTTGTGAGAACATGCCAGAGTGGAGCA 3276  
 QY 1701 SerGluValAspLysGlnLeulleGluArgCysSerArg----- 1713

Db 3277 ATTTCTAAATATTTGTTTGTATCCGAAAGTTGTAATCGCAGTATTCCTCTGATGGAAGTC 3336  
 QY 1714 -----llelleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsn 1727  
 Db 3337 ATCAGATATGCTGCAAGTCTTGTCTTAATGTATCTTAAGTATGAGAAAACACTACTTCAGCA 3396  
 QY 1728 ThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
 Db 3397 GTTTATGATGTAGAAAATGTATAGATATACTATTGGAGCTTTTGCAGATATACCGAGAA 3456  
 QY 1743 -----ArgTrpCysAspLysAspSerGluillePheAsnThrLeuCysThrLeu 1758  
 Db 3457 AAGCCTGGTAAATAAGATTGCAGACAAAGCGGAGCATTTTACAAAAACTTGTGTGTTG 3516  
 QY 1759 IleTrpValPheAlaHisCysProLysLysArgLysIlelleHisAspTyrMetThrAsn 1778  
 Db 3517 TTGGCTATTTTACTG-----AAGACAACAAAATAGAGCCTCTGATGTACGAAGTAGG 3567  
 QY 1779 ProGluAlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMet 1798  
 Db 3568 TCCAAAGTTGTTGACCGTATTTACAGTCTCTACAACTTACAGCTCATAAACATAAATG 3627  
 QY 1799 LysGlnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys--- 1817  
 Db 3628 ---AATACTGAAAGATA-----CTTTACAAGCAAAAGAGAAT 3663  
 QY 1818 -----lleAsnPheThrProCysSer-----LeuProSer 1827  
 Db 3664 TCTTCTATAGCAATCTCTTTTATCCCAAGAAACACCTGTAGGACCAGAAATAGTTTCAAGA 3723  
 QY 1828 LeuGluProAspPheGlyllelleArgTyrSerProTyrThrPheIleSerSerValTyr 1847  
 Db 3724 CTTAAGCCAGATTGGGTTTGGAGACAGATAACATGAGAAATACCAAAATCCCTGCCAA 3783  
 QY 1848 AlaPheAspThrIleLeuLysLysLeuGlnIle 1858  
 Db 3784 GCTATTCAAATGGTGTGATGATACGCTTGCAAT 3816  
 RESULT 6  
 ABQ99318  
 ID ABQ99318 standard; cDNA; 4027 BP.  
 XX  
 AC ABQ99318;  
 DX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human coding sequence SEQ ID 51.  
 XX  
 KW Human; expressed sequence tag; EST; chromosome 17;  
 KW haematopoietic disorder; central nervous system disease; viral infection;  
 KW peripheral nervous system disease; non-healing wound; infectious disease;  
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;  
 KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;  
 KW antiallergic; antiinflammatory; immunosuppressive; fungicide;  
 KW cytostatic; haemostatic; virucide; antibacterial; neuroprotective;  
 KW immunostimulant; cerebroprotective; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200259260-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 16-NOV-2001; 2001WO-US042950.  
 XX  
 PR 17-NOV-2000; 2000US-00714936.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Goodrich RM, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;  
 PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 XX

DR WPI: 2002-590824/63.  
 DR N-PSDB; ABP64732.  
 XX  
 PT New isolated polynucleotide, useful in research, diagnostic or  
 PT therapeutic methods, e.g. preventing or treating disorders involving  
 PT aberrant protein expression or biological activity.  
 XX  
 PS Claim 1; SEQ ID NO 51; 394pp; English.  
 XX  
 CC The present invention relates to novel human coding sequences (ABQ99268-  
 CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in  
 CC therapeutic, diagnostic and research methods. The polynucleotides may be  
 CC used in the field of molecular biology as hybridisation probes, primers  
 CC for PCR, for chromosome and gene mapping, for the recombinant production  
 CC of protein, or in generation of anti-sense DNA or RNA. The  
 CC polynucleotides are useful in diagnostics as expressed sequence tags  
 CC (ESTs) for identifying expressed genes or for physical mapping of the  
 CC human genome. The proteins may be used as molecular weight markers, or as  
 CC nutritional sources or supplements. The proteins may be used to maintain  
 CC and expand cell population in a totipotent or pluripotent state  
 CC useful for re-engineering damaged or diseased tissues, transplantation,  
 CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
 CC polynucleotides and proteins are useful for preventing, treating or  
 CC ameliorating disorders involving aberrant protein expression or  
 CC biological activity, e.g. haematopoietic disorders, central/peripheral  
 CC nervous system diseases, mechanical and traumatic disorders, non-healing  
 CC wounds, immune deficiencies and disorders, infectious diseases caused by  
 CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
 CC reactions and conditions, coagulation disorders, or cancer. The  
 CC polynucleotide sequences of the invention were assembled from ESTs  
 CC isolated mainly by sequencing by hybridisation, and in some cases,  
 CC sequences obtained from one or more public databases. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic form directly from WIFO at  
 CC ftp.wifo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 4027 BP; 1492 A; 672 C; 770 G; 1093 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6,9e-41 Length: 4027  
 Score: 700.50 Matches: 275  
 Percent Similarity: 40.70% Conservative: 226  
 Best Local Similarity: 22.34% Mismatches: 412  
 Query Match: 7.36% Indels: 319  
 Ds: 6 Gaps: 40

US-09-914-698-1 (1-1861) x ABQ99318 (1-4027)

QY 868 TrpGlnLeuIleTyrIlysPheArgSerProLysPheHisAlaAlaThrValLeuGln 887  
 Db 359 TGGATGATTAAGAAAGGATGCGA-----GAGATGCCAGGGCTGCTACTTTCATCCAG 412  
 QY 888 LysTrpTrpArgArgHisTrpLeuHis-----Val 897  
 Db 413 TCTACTTTCAGATGACAGATTAACATATACAGATATCAGGCTTTGAAACAGGGCTCCGTT 472  
 QY 898 ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet 909  
 Db 473 GTGATCCACACGCAATACCAAGCAATAGAGTGCAGAACTGCAGAGCGCATTTATCTC 532  
 QY 910 ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys 929  
 Db 533 AGCAAAAGACACTCTGCTGTGATCCTTCAGGCTGCATTTCAGGGGTATGAAACTAGAGA 592  
 QY 930 TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThr 949  
 Db 593 CATTTGAGAGTATGCATTCC-----TCTGCAACCTTATTCAGAGTAGGTTT 640  
 QY 950 ArgArgTyrLeuAlaGlnLysGlnLeuTyrHisSerIleIleThrIleGln 969  
 Db 641 AGATCATTAATGTTGGTGGAGGAGAGATTTCATTTCCCTCAAAAAGCTACTATTTTGTTCAG 700  
 QY 970 ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg 989

Db 701 AGGAATATCGAGCCACCATTGTCGCAACATAAATTTGTACCAATTTCTGCACTTAAGA 760  
 QY 990 GluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeu--- 1008  
 Db 761 AAGGAGCCATTACAAATCAGTCATCTTACAGAGACTGATGTTAAAGAAAGAGTATCAA 820  
 QY 1009 -----LeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln--- 1023  
 Db 821 GAATGCAAAAGGGCTGCAGT-TCTCATTCAGGCTACTTTTCAGGATGCACAGAAAAA 879  
 QY 1024 -----GlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThr 1036  
 Db 880 TATATTACATTTTCAGACTTGGAAACATGCTTCAATTTCAATTCAGCAACATTATCAACA 939  
 QY 1037 TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe 1056  
 Db 940 TATAGAGCTGCAAAATTCGAAGAGAAATATATATACAGCAATGCGATTCTCTCTGGTT 999  
 QY 1057 ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGlu 1073  
 Db 1000 ATTGAGCTGCATATATAAGGAATGAAGCAGACAACTTTTAAGGGAAGAAACACAAAGCT 1059  
 QY 1074 -----PheGlnGlnLeu 1077  
 Db 1060 TCTATCGTAATACAAAGCACCTTACAGAAATGTATAGGCAGTATTTGTTCTACCAAAAGCTT 1119  
 QY 1078 LysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLysCys 1097  
 Db 1120 CAGTGGCTTACAAAATCATAACAGAAATATATAGCAGCAATATAAAGAAAGAAACAGAAA 1176  
 QY 1098 AsnAlaAspTyrLeuAlaLeuArgSerValLeuLys-----ValGlnAla 1113  
 Db 1177 -----GTATTTCAACATGAACTTAAGAAAGAGACTTGTGTTGAGGCA 1221  
 QY 1114 ---TyrArgLysAlaThrIleGlnMetArgIleAsp----- 1124  
 Db 1222 GGTTCAGGACATGCAACATAAAAAACAGATTCAGGAACAGCACCGGCTGCCATTATT 1281  
 QY 1125 -----ArgAsnHisTyrTyrSerLeuArgLysAsn 1134  
 Db 1282 ATTCAGAAAGCAATGTAAAGCCTTTAAATAAGGAAGCAATATCTCCACCTTAGAGCAACA 1341  
 QY 1135 ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGlnGln----- 1151  
 Db 1342 GTAGTTTCTATTCAAGAGATACAGAAACTAACTGTCAGTCGCTACCCAGCAGTTATT 1401  
 QY 1152 -----ArgGluAsnTyrLeuArgLeuArgAsn 1160  
 Db 1402 TGTATACAGTCTTATTACAGAGGCTTTAAAGTACGAAAGGATATTCAAAATATGACCGG 1461  
 QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
 Db 1462 GCTGCCACATTAATCAGTCATCTATCGAATGCACAGGGCCAAAGTTGAT----- 1512  
 QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
 Db 1513 ---TATGAACAAAGAAACTCAATTTGGTGTATACAGAAATATATAGTTGTATGTT 1569  
 QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln 1220  
 Db 1570 AGAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1629  
 QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLys 1240  
 Db 1630 GCTGCTTTTAGAGGCGATGAAA-----GTAGACAAAAATTCGAAAAATGATCAGAGGAA 1683  
 QY 1241 LysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 Db 1684 AGATGGCGCAGCCATTGTT-----AACCAATTCGCACTCTGCTGTTACAGAACCAAA 1734  
 QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280



Qy 1848 AlaPheAspThrIleLeuCysLysLeuGlnIle 1858  
 Db 3784 CTTATTCAAATGGTGTGATGATACGCTTGGCATT 3816

## RESULT 7

ABV20287/c  
 ID ABV20287 standard; cDNA; 1115 BP.

XX AC ABV20287;  
 XX AC  
 XX DT 13-SEP-2002 (first entry)  
 XX DE  
 XX DE Human prostate expression marker cDNA 20278.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PS Claim 1; Page 3317-3318; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservatives: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV20287 (1-1115)

Qy 638 lIeLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValVal 657  
 Db 1092 GTTCCTGAATTGGCTGTGTTCCTACAAATCCTTTGGCTTCGAATTGGCTACAGACAACT 1033

Qy 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db 1032 TATGGAGAACTCATATCTTTGGAGATACAGTGTATGTCACAGGCTTGGCTATGTTATT 973  
 Qy 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTTGGAAATCCTGATATAGCAGCTGAGTATAGACACCCACTGTTCT 913  
 Qy 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 Db 912 CACCTGTATAGAGATGGTCATGAAGAGCTTTGTCCAACTTTACATGAAAAGTTATG 853  
 Qy 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 Db 852 TTGTGGTCTGTTTCTTGTATATGCTAAATTCAGACTCATGTGATCATGATCTTGT 793  
 Qy 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCCTCTGTAAGATGCGCAATTCAGGCTAGTAAAGAAATCCTTTTGGCTTTTTCACGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGGTGAAGGTGACCTTTCCCGTCACCTTGGCTTATGGGATTACCTGT 673  
 Qy 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTTACAGACACCATTTGATGAATTTGATTTGGCGTTACAAATCTTGGCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGGAGTGGCTTGTGCGAACCATGGAACTTCTCACACAGAACTGGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAAACTCAGGATTCGGCAATAAGTCTCTTCAAGATGCAACAATGTCACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGAGATTTGAATTAAGTATGATGATGAGCAATAACAATT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuTrpGlnLeu 870  
 Db 432 CTATCTAAGGATATTGTGGATAGGCACAGAGAAAAAACTCTCAGATTGCTTGGAAAAATA 373  
 Qy 871 lIeTyrLysPheArg 875  
 Db 372 GCGTTTGTCTTTTCAG 358  
 RESULT 8  
 ID ABV20247/c  
 XX ABV20247 standard; cDNA; 1115 BP.  
 XX AC ABV20247;  
 XX DT 13-SEP-2002 (first entry)  
 XX DE Human prostate expression marker cDNA 20238.  
 XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 XX KW pharmacogenomic marker; gene; ss.  
 XX OS Homo sapiens.  
 XX PN WO200160860-A2.  
 XX PD 23-AUG-2001.  
 XX PF 20-FEB-2001; 2001WO-US005171.  
 XX PR 17-FEB-2000; 2000US-0183319P.  
 XX PR 16-MAR-2000; 2000US-0189862P.  
 XX PR 25-MAY-2000; 2000US-0207454P.  
 XX PR 09-JUN-2000; 2000US-0211314P.



Alignment Scores:  
 Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV26078 (1-1115)

QY 638 lleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuValVal 657  
 DB 1092 GTCTGAATGGCTGTGTCCTACATCTTTGGTTCGAATGGCTAGAGCAACT 1033  
 QY 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 DB 1032 TATGGAGAACTCATATCTTTGGAAGATAACAGTGATGTCCAGGGTTGGCTATGTTATT 973  
 QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 DB 972 CTGAATCGCTACTTTGGAAATCCTGATATAGCAGCTGATATAGACACCCCACTGTCTCT 913  
 QY 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 DB 912 CACCTGTATAGAGATGGTCATGAAGAGCTTTGTCCAGTTTACATTTGAAGAAGTATTG 853  
 QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 DB 852 TTGTGGTCTGTCTTTCTGATGATGCTAAATTTCCAGACTCATTTGATCATGCTCTGT 793  
 QY 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 DB 792 CTCCTTGTAAAGATCCGGAATTCAGGCTAGTAAGAAATCTTTGGCTTTTTCACGA 733  
 QY 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeu 774  
 DB 732 GATTTCTTAAGTGGTGAAGTGACCTTTCCCGTCACCTTGGCTTATTTGGATTAAGTCT 673  
 QY 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 DB 672 AACCATGTTTCAGACACCAATTTGATGAATTTGATTTTGGCGTTACAAATCTTCCGTAG 613  
 QY 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 DB 612 TTGCAATGTGGAGTGGCTTTGTCCGAACCATGGAACCTTCACACAGAACTGGGACCTC 553  
 QY 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 DB 552 TCAAGAAACTCAGGATTCGGCAATAAGTCGTCTTCAAAGATGCACAAATGTGACATT 493  
 QY 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 DB 492 GTTCTTCAAGTCTTAAATCATCGAGGAATTTGAATTAAGTGATGAGCATGGAATAACA 433  
 QY 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 DB 432 CTATCTAAGGATATTGTGATAGGCACAGAGAAAACCTCTCAGATTGCTTTGGAAATA 373  
 QY 871 IleTyrLysPheArg 875  
 DB 372 GCGTTTGTCTTTTCAG 358

RESULT 10  
 ID ABV20220/c  
 ABV20220 standard; cDNA; 1115 BP.  
 XX  
 AC ABV20220;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 20211.  
 XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 XX pharmacogenomic marker; gene; ss.

OS Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX 16-MAR-2000; 2000US-0189862P.

XX 25-MAY-2000; 2000US-0207454P.

XX 09-JUN-2000; 2000US-0211314P.

XX 18-JUL-2000; 2000US-0219007P.

XX 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Claim 1; Page 3304; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient

XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV20220 (1-1115)

QY 638 lleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuValVal 657  
 DB 1092 GTCTGAATGGCTGTGTCCTACATCTTTGGTTCGAATGGCTAGAGCAACT 1033

QY 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 DB 1032 TATGGAGAACTCATATCTTTGGAAGATAACAGTGATGTCCAGGGTTGGCTATGTTATT 973

QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 DB 972 CTGAATCGCTACTTTGGAAATCCTGATATAGCAGCTGATATAGACACCCCACTGTCTCT 913

QY 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 DB 912 CACCTGTATAGAGATGGTCATGAAGAGCTTTGTCCAGTTTACATTTGAAGAAGTATTG 853

QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 DB 852 TTGTGGTCTGTCTTTCTGATGATGCTAAATTTCCAGACTCATTTGATCATGCTCTGT 793

Db 852 TTGTTGGCTGCTTTCTTGTATGATGCTAAATTTCCAGACTCATTTGATCATGATCCTTGT 793  
 Qy LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCTCTGTAAGATGCGAATTCAGGCTAGTAAAGAAATCCTTTGGCTTTTTCACGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLysGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGTGAAGTGACCTTTCCCGTCACCTTGGCTTATTGGGATTACCTGTT 673  
 Qy 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTCAGACACCATTTGATGAATTTGCTTTCCCGTTACAAATCTTTGCCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGAGTGGCTTGTGGACCATGGAATTCACAGAACTGGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAACTCAGGATTCGGCAATAGTCTCTTCAAAAGATGCACAAATGTTGACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGAAATTCGAATTAAGTGTGATGATGCAATACCAATT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTATCTAAGGATATGTGGATAGGCACAGAGAAATACTCTCAGATTGCTTTGGAAATA 373  
 Qy 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGGCTTTTCAG 358  
 RESULT 11  
 ID ABV26049/c  
 XX ABV26049 standard; cDNA; 1115 BP.  
 AC ABV26049;  
 DX  
 TT  
 XX 16-SEP-2002 (first entry)  
 DE Human prostate expression marker cDNA 26040.  
 XX  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 5269; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX  
 SQ Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. NO.: 4.06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2  
 US-09-914-698-1 (1-1861) x ABV26049 (1-1115)  
 Qy 638 IleLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuValVal 657  
 Db 1092 GTCTGAAATGGCTGTTGTTCTCAATCCTTTGGCTTCGAATGGTCTAGACAACT 1033  
 Qy 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db 1032 TATGGAGAACTCATATCTTTGGAGAGATAACAGTGTGTACAGGGTTGGCTATCTTATT 973  
 Qy 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTTGGAAATCTGTATATGATGATGATGATGATGATGATGATGAT 913  
 Qy 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 Db 912 CACCTGTATAGATGCTGTATGAAAGCTTTGCCAAGTTTACATTTGAAATAAATTATG 853  
 Qy 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 Db 852 TTGTTGGTCTGTTTCTTGTATGATGCTTAAATTTCCAGACTCATTTGATCATGATCCTTGT 793  
 Qy 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCTTCTGTAAGATGCGAATTCAGGCTAGTAAAGAAATCCTTTTGGCTTTTTCACGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGTGAGTGGCTTGTCCCGTCACCTTGGCTTATTGGGATTACCTGTT 673  
 Qy 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTTCCAGACACCATTTGATGAATTTGATTTTCCCGTTTCAAAATCTTTGCCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGAGTGGCTTGTGGCAATTCAGGAACTTCTCACAGAACTGGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAACTCAGGATTCGGCAATAGTCTCTTCAAAAGATGCACAAATGTTGACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGAAATTCGAATTAAGTGTGATGATGCAATACCAATT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTATCTAAGGATATGTGGATAGGCACAGAGAAATACTCTCAGATTGCTTTGGAAATA 373



QY 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGCTTTTCAG 358  
 RESULT 12  
 ID ABV26118/c  
 XX ABV26118 standard; cDNA; 1115 BP.  
 AC ABV26118;  
 XX  
 XX 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 26109.  
 XX  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200160860-A2.  
 XX  
 XX 23-AUG-2001.  
 XX  
 XX 20-FEB-2001; 2001WO-US005171.  
 XX  
 XX 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 XX Schlegel R, Endege WO, Monahan JE;  
 PI  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 XX Claim 1; Page 5283; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX  
 XX Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2  
 US-09-914-698-1 (1-1861) x ABV26118 (1-1115)  
 QY 638 lleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGluValVal 657  
 Db 1092 GTCCCTGAATTGGCTGTGTCCTCAACATCTTTGGCTCGAATTTGGCTAGAGCAACT 1033

QY 658 PheGlyGluLysIleGlnMetClnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db 1032 TATGAGAACTCATATCTTTGGAGATAACAGTGAATGTCACAGGGTTGGCTATGTTATT 973  
 QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTTGGAAATCCTGATATAGCAGCTAGTATAGACACCCACTGTTCT 913  
 QY 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysLeu 714  
 Db 912 CACCTGTATAGAGATGCTCATGAAGAAGCTTGTCCAAAGTTTACATGAAAAAGTTATTG 853  
 QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisnProCys 734  
 Db 852 TTGTGGTCTGTCTTTCTTGAATATCTAAATTTCCAGACTCATTGATCATCATCTTGT 793  
 QY 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCCTCTGAAGATGCCGAATTCAGGCTAGTAAGAAATCCTTTTGGCTTTTTCACGA 733  
 QY 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGTGAAGGTGACCTTTCCCGTCACTTGGCTTATTTGGATTACCTGT 673  
 QY 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeuAlaValAsp 794  
 Db 672 AACCATGTCAGACACCAATTTGATGAATTTGCGTTTACAAATCTTTCGCTAGAC 613  
 QY 795 LeuArgAspGlyValArgLeuThrArgValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGGAGTGGCTTGTGGCAACCATGGAACTTCTCACACAGACTGGGACCTC 553  
 QY 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAAACTCAGGATTCGCGCAATAAGTCGCTCTTCAAAAGATGCACAATCTGCATT 493  
 QY 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTTCTTAAATCAGGAATTAAGTAAAGTATGATGATGATGATGATGAT 433  
 QY 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTAICTAAGGATATTTGTGATAGGCACAGAGAAATACTCTCAGATTGCTTTGGAATA 373  
 QY 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGCTTTTCAG 358  
 RESULT 13  
 ABV09513  
 ID ABV09513 standard; cDNA; 752 BP.  
 AC ABV09513;  
 XX  
 XX 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 9504.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200160860-A2.  
 XX  
 XX 23-AUG-2001.  
 XX  
 XX 20-FEB-2001; 2001WO-US005171.  
 XX  
 XX 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.





```
QY 1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543
Db 1247 TTTACAGAAAGAGATTATTTCAGAAATATCATAGCATC-----1285
QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563
Db 1286 -----AAAAGATTGAGCATGAAGTCAAGAATGCTGTGAGCCAGCA 1327
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583
Db 1328 AATAGGCGTCATCAGTAATACAGAAAGCAGTCGCCAATTTCTCTCCGTAAAGACAG 1387
QY 1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysVal 1603
Db 1388 GAAAAATTCATAGT-----GGAATCATTA-----1414
QY 1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGlnGlyLeuLeu 1623
Db 1415 ATTACGGCATTTATGAGAGGCTATTCTTGAGGAAGAAATAATGATTGTACAAAAATTAAA 1474
QY 1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg 1643
Db 1475 GCTATACGACTAAGTCTTCAAGTTGTTAATAGGAGATTTCGAGAAAGAAACAACCTCTAC 1534
QY 1644 CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663
Db 1535 AAAAGAACTGCACCTGCATTCATTACCTTTTGACATATAAGCACCTTCTGCCATTCCT 1594
QY 1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1683
Db 1595 GAGGCGCTTAAACACCTAGAGGTAGTTACTAGA-----1627
QY 1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703
Db 1628 -----TTGTCTCCACTTTGTGTGAGAAACATGCCCGAGAGTGAGCAATTTCTAAA 1678
QY 1704 AspLysGlnLeuIleGluArgCysSerArg-----1713
Db 1679 ATATTGTTTTCGGAAGTTGTAATCGCAGTATTCCTTGATGGAAGTCATCAGATAT 1738
QY 1714 -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730
Db 1739 GCTGTGCAAGTCTGTGTTATGATCTAAGTATGAGAAACTACTTCAGCAGTTTATGAT 1798
QY 1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu-----1742
Db 1799 GTAGAAAATTGTATGATATCTATTGAGAGCTTTTGACAGATATACGAGAAAACCCCTGGT 1858
QY 1743 ---ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVal 1761
Db 1859 AATAAAGTTGCAGACAAAGCGGAAGCATTTTACAAAACTTGTGTTGTTGGCTATT 1918
QY 1762 PheAlaHisCysProLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781
Db 1919 TTAAGTCTG-----AAGACAAACAAATAGAGCCCTCTGATGACGAGTAGGTCCAAAGTT 1969
QY 1782 IleTyrMetValArgGluThrLysLeuValAlaArgLysGluLysGlnAsn 1801
Db 1970 GTTGACGGTATTACAGTCTCTACAACTTACAGCTCATAAACATAAATG-----AAT 2023
QY 1802 AlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys-----1817
Db 2024 ACTGAAAGATA-----CTTTACAAAGCAAAAGAAAGAAATTTCTCTATA 2065
QY 1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
Db 2066 AGCATTTCTTTTATCCAGAAACACCTGTAAGGACCAAGATAGTTTCAAGACATTAAGCCA 2125
QY 1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerValTyrAlaPheAsp 1850
Db 2126 GATGGGTTTGTAGAGAGATACATGAGAGAAATCACAAATCCCTGCAAGCTATTCAA 2185
QY 1851 ThrIleLeuCysLysLeuGlnIle 1858
```

```
Db 2186 ATGGTGTATGATACGCTTGGCAT 2209
      ::::: ||| |||
```

## RESULT 15

ACC50994

ID ACC50994 standard; cDNA; 2414 BP.

XX ACC50994;

AC ACC50994;

XX 12-JUN-2003 (first entry)

XX Human bladder cancer associated cDNA sequence SEQ ID NO:81.

XX Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.

XX Homo sapiens.

XX W02003003906-A2.

XX 16-JAN-2003.

XX 03-JUL-2002; 2002WO-US021338.

XX 03-JUL-2001; 2001US-0302814P.

XX 03-AUG-2001; 2001US-0310099P.

XX 08-NOV-2001; 2001US-0343705P.

XX 13-NOV-2001; 2001US-0350666P.

XX 12-APR-2002; 2002US-0372246P.

XX (BOSB-) EOS BIOTECHNOLOGY INC.

XX Mack DH, Aziz N;

XX WPI; 2003-201532/19.

XX P-PSDB; ABR48183.

XX Detecting a bladder cancer-associated transcript in a cell from a patient, comprises contacting a biological sample from the patient with a bladder cancer-associated polynucleotide or antibody.

XX Claim 6; Page 261; 307pp; English.

XX The present invention describes a method for detecting a bladder cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with a polynucleotide that selectively hybridizes to a sequence that is 80 % identical to a table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059 encode the human bladder cancer-associated proteins given in ABR48146 to ABR48242). Bladder cancer-associated sequences from the present invention have cytostatic activities, and can be used in antisense gene therapy and in vaccine production. The method can be used for detecting a bladder cancer-associated transcript in a cell from a patient. The method is useful in diagnosing or treating bladder cancer and in screening for compounds that modulate bladder cancer, such as hormones or antibodies. The nucleic acid molecules from the present invention may be used in various screening and diagnostic methods, and for gene therapy, vaccine and/or antisense/inhibition applications

XX Sequence 2414 BP; 859 A; 391 C; 476 G; 688 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	5,97e-22	Length:	2414
Score:	437.00	Matches:	210
Percent Similarity:	36.74%	Conservative:	153
Best Local Similarity:	21.26%	Mismatches:	331
Query Match:	4.59%	Indels:	294
DB:	7	Gaps:	33

US-09-914-698-1 (1-1861) x ACC50994 (1-2414)

QY 913 ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrVallys 932

Db 2 AAATCTGTACGAACACTATTCAGGCTGCTTTTAGAGGCATGAAGTTAGACAAAATTGAAA 61

933 LeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyr 952  
Db : : : : :  
62 AATGATATCAGAGGAAAG---ATGCGACCCATTGTTAACCAATTCGCACTCTGCTGTATC 118  
Qy : : : : :  
953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr 971  
Db : : : : :  
119 AGAAGCAAACTCAG---TATGAGCTGTTCAAGTGNAGGTGTATGATTCAGAGTGG 175  
Qy : : : : :  
972 TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla 991  
Db : : : : :  
176 TATAAAGCTTCTGCTTGTGTTTCACAGCAAGCAGAGTATCATCTCTCAAAAGTAGGCT 235  
Qy : : : : :  
992 AlaIlePheLeuGlnArgIleTrpArgArgGlnPheAlaLysLysLeuLeuAla 1011  
Db : : : : :  
236 GCAGTAACAATTCAAAAGCTTTT---TGTAGAATGGTCACAGAAACTG--- 283  
Qy : : : : :  
1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAspTyrIleGln 1031  
Db : : : : :  
284 ---GAAACACAGAAATGTGCTCCCTACGGATTTCAG 316  
Qy : : : : :  
1032 ---MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg 1049  
Db : : : : :  
317 TTCTTCCTTCAGTGGCTGTAT---CGGAGAAATTTGTTTCAG 358  
Qy : : : : :  
1050 GlnArgAspLeuIleMetPheValGlnArgMetArgSerLysTyrSerMetLeuGlu 1069  
Db : : : : :  
359 CAGAAAAGAGCTGCTCACTTACAGCATTTATTTAGACG---TGG---CAA 406  
Qy : : : : :  
1070 GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnArgTyrArg 1089  
Db : : : : :  
407 ACCGAAACAGT---TTTACTATATAGAAAGCAGCGTGGTTTACAAATCACTACAGA 466  
Qy : : : : :  
1090 AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu 1109  
Db : : : : :  
467 GCATTTCCTGTCGCAAAACATCAAGACAGAGCTATTTACAGATCAGAAAGCAGTGTATC 526  
Qy : : : : :  
1110 LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyr 1129  
Db : : : : :  
527 ATTATTCAGCTAGAGTAAAGGATTTATACAGAAACG--- 565  
Qy : : : : :  
1130 SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg 1149  
Db : : : : :  
565 --- 565  
Qy : : : : :  
1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln 1166  
Db : : : : :  
566 ---AAGTTTCAGGAAATTAATAATAGCACCATTAATAAATTCAGGCTATGTGG 613  
Qy : : : : :  
1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186  
Db : : : : :  
614 AGGAGATATAGACCCCAAGAAA---TATTTATGTAAGTG 649  
Qy : : : : :  
1187 LysCysIleIleAsnValGln---ArgArgTyrArgAlaThrLeuGlnMetArg 1203  
Db : : : : :  
650 AAAGCTGCTGCAGAAATTCAGCTGTATAGATGTTGGAGAGCA--- 694  
Qy : : : : :  
1204 ArgGluArgLysAsnTyrLeuHisLeuThrThrThrLysArgIleGlnLysPhe 1223  
Db : : : : :  
695 ---CACAAAGAAATATCTAGCTGTATTAAAGCTGTATAAATATTCAAGTTGCTTC 748  
Qy : : : : :  
1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThr 1243  
Db : : : : :  
749 TATACCAAACTAGAG---AGAACACGCGTTTTCGAATGTGAGACATCAGCA 796  
Qy : : : : :  
1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263  
Db : : : : :  
797 ATTATCAITTCAGAGAAATGGAGAGCTATCTCTCGCAAGAGTAGCTCATGACACTTC 856  
Qy : : : : :  
1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMet 1283  
Db : : : : :  
857 TTAATGATAAAAGA--- 871

1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis 1303  
Db : : : : :  
872 ---CATCGAGCTGCTTGTGTTGATCCAGCACAT 901  
Qy : : : : :  
1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323  
Db : : : : :  
902 TATAGAGCA--- 910  
Qy : : : : :  
1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343  
Db : : : : :  
911 ---TATAAGGAAGGAGGCTCTCTCTTCGGCAG--- 940  
Qy : : : : :  
1344 TyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363  
Db : : : : :  
941 ---AAATCTGCTGCTTTGATCATACAAAAATATATACGAGCAGGAGGCT 988  
Qy : : : : :  
1364 MetGlnLysGlnValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArg 1383  
Db : : : : :  
989 GCAAAGCATGAAGGATAAAATATATTGAATTTAAAAAATCTACAGTTATCTACAGCA 1048  
Qy : : : : :  
1384 ArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGln 1403  
Db : : : : :  
1049 CTGTGCTGCTGCTGCTAGTAGTACGAAAGATTTTAGAA---CAGAGA 1093  
Qy : : : : :  
1404 AlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeuCysLysAla 1423  
Db : : : : :  
1094 GCCAAATTCACCT---CTTCACCTTCACCTGACGCT 1126  
Qy : : : : :  
1424 GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrVal 1443  
Db : : : : :  
1127 GCATATTATACCTGAAT--- 1144  
Qy : : : : :  
1444 GlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
Db : : : : :  
1144 --- 1144  
Qy : : : : :  
1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArg 1483  
Db : : : : :  
1144 --- 1144  
Qy : : : : :  
1484 ArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db : : : : :  
1145 ---GCTGTTAGAAATTCAAAGAGCCTATAAATTTACCTGGCTGTGAAGAATGCTAAC 1198  
Qy : : : : :  
1504 GlnAspTyrGlnLeuLeuGlnSerSerValIleLeuValGlnArgCysPheArgAlaAsn 1523  
Db : : : : :  
1199 AAGCAGGTT---AATTCAGTCACTCTGTTTACAGAGATGTTTCGAGCAAGA 1246  
Qy : : : : :  
1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db : : : : :  
1247 TTACAAGAAAGAGATTTTATTCAGAAATATCATAGCATC--- 1285  
Qy : : : : :  
1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db : : : : :  
1286 ---AAAAAGATTGAGCATGAAGGTCAAGAAATGTCGTGAGCCAGCA 1327  
Qy : : : : :  
1564 ArgCysSerSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db : : : : :  
1328 AATAGGCTGCATCAGTAATACAGAAAGCAGTGCCCATTTCTCTCGTAAAGCAG 1387  
Qy : : : : :  
1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaLysVal 1603  
Db : : : : :  
1388 GAAAAATTCACCTAGT---GGAATCATTA--- 1414  
Qy : : : : :  
1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGlnGlyLeuLeu 1623  
Db : : : : :  
1415 ATTCAAGCATTTATGAGAGGCTATTCTTGGAGAAAGAAATGATGTACAAAAATTA 1474  
Qy : : : : :  
1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg 1643  
Db : : : : :  
1475 GCTATACGACTAAGTCTTCAAGTTGTTAATGGGAGATTCGAGAGAAACAACTCTAC 1534  
Qy : : : : :  
1644 CysLysLysGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663

```
Db      :::      |||:::  :::|||      |||      |||      |||
1535 AAAGAACTGCACCTTGCACTTCATTACCTTTTGACATATAAGCACCTTTCTGCCATTCTT 1594
Qy      |||
1664 AlaValLeuSerGlnLeuSerArgThrValProHisLeuLeuMetTipCys 1683
Db      |||      |||:::  |||
1595 GAGCCTTAAACACCTAGAGTAGTTACTAGA----- 1627
Qy      |||
1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703
Db      :|||
1628 -----TTGTCCTCCACTTTGTTGTGAGAACATGGCCAGAGTGGAGCAATTTCTAAA 1678
Qy      |||
1704 AspLysGlnLeuIleGluArgCysSerArg----- 1713
Db      |||
1679 ATATTGTTTGTATCCGAAGTGTAAATCGCAGTATTCCTTGTATGGAAGTCATCAGATAT 1738
Qy      -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730
Db      :|||
1739 GCTGTGCAAGCTCTGCTTAAGTATCTAAGTATGAGAAACTACTTCAGCAGTTTAIGAT 1798
Qy      |||
1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742
Db      :|||
1799 GTAGAAAATTGTATAGATATACTATTGGAGCTTTTGCAGATATACCGAGAAAAGCCTGGT 1858
Qy      ---ArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyrVal 1761
Db      :|||
1859 AATAAAGTTGCAGACAAAGCGGAGCATTTTACAAAACTTGTGTTTGTGGCTATT 1918
Qy      |||
1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781
Db      :|||
1919 TTACTG-----AAGACACAAATAGAGCCTCTGATGTACGAAGTAGTCCAAAGTT 1969
Qy      |||
1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801
Db      :|||
1970 GTTGACCGTATTACAGCTCTCAAACTTACAGCTCATAAACATAAAATG-----AAT 2023
Qy      |||
1802 AlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817
Db      :|||
2024 ACTGAAGAATA-----CTTTACAAGCAAAAGAAATTTCTCTATA 2065
Qy      ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
Db      |||
2066 AGCATTCCTTTTATCCAGAAACACCTGTAAAGCAGACAGATAGTTTCAAGACTTAAGCCA 2125
Qy      |||
1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
Db      |||
2126 GATTGGGTTTGTGAGAAGAGATAACATGGAAGAAATCACAATCCCTGCAAGCTATTCAA 2185
Qy      |||
1851 ThrIleLeuCysLysLeuGlnIle 1858
Db      :|||
2186 ATGGTGATGATACGCTTGCAATT 2209
```

Search completed: September 15, 2004, 02:07:39  
Job time : 1632 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 15, 2004, 00:52:24 ; Search time 9578 Seconds  
(without alignments)  
5802.205 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool\_h/US09914698/runat\_10092004\_150005\_28235/app\_query.fasta\_1.2055  
-DE=EST -QFMT=tastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09914698 @CGN 1.1 7670 -runat\_10092004\_150005\_28235 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rtd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1391	14.6	1101	29	CNS00G5G
2	1115	11.7	1022	29	CNS00LY6
3	1002	10.5	596	9	AI519740
4	983.5	10.3	911	29	CNS015NX
5	851	8.9	718	12	BI642052
6	723.5	7.6	3130	11	BC026699
7	716.5	7.5	3983	11	AK052326
8	647	6.8	589	12	BG636223
9	475	5.0	2675	11	AK048277
10	429	4.5	485	28	BH374743
11	388	4.1	1008	29	CNS04QCR
12	359.5	3.8	526	9	AL720781
13	359.5	3.8	526	9	AL720782
14	334	3.5	619	28	CC129703
15	329	3.5	852	29	CNS046XE
16	316	3.3	771	9	AUI27218
17	309	3.2	983	13	BUI93761
18	304.5	3.2	755	13	BUI222936
19	291.5	3.1	789	13	EX077268
20	287	3.0	919	13	BQ952955
21	285	3.0	582	13	BUI206367
22	282.5	3.0	420	14	CD036256
23	280.5	2.9	689	12	BGI77316
24	278.5	2.9	937	13	EX702765
25	277	2.9	592	9	AA539252
26	273	2.9	792	13	BUI055628
27	270	2.8	839	13	BUI420251
28	269.5	2.8	696	13	EX087585
29	269	2.8	771	9	AL718233
30	269	2.8	772	13	BM079393
31	268.5	2.8	773	14	CA381839
32	267	2.8	704	12	BGI75972
33	267	2.8	725	14	CD350820
34	267	2.8	1131	13	EX383325
35	265	2.8	769	13	EX077267
36	263	2.8	856	13	BU911386
37	262.5	2.8	664	28	AZ384962
38	257.5	2.7	819	12	BGI14414
39	255.5	2.7	4137	29	AY420057
40	253.5	2.7	955	29	CNS046XF
41	253.5	2.7	6294	29	AY410631
42	252	2.6	566	14	CF178179
43	251.5	2.6	711	9	AL718189
44	249.5	2.6	6237	29	AY410633
45	248.5	2.6	829	14	CF289447

# ALIGNMENTS

RESULT 1  
CNS00G5G  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR32J06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL071458  
VERSION AL071458.1 GI:4951500  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;



Ephydroidea: Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCL-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

FEATURES  
source

1. 1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BACR32J06"  
/note="end : TET3"

ORIGIN

Alignment Scores:  
Pred. No.: 1.94e-108 Length: 1101  
Score: 1391.00 Matches: 280  
Percent Similarity: 92.46% Conservative: 2  
Best Local Similarity: 91.80% Mismatches: 3  
Query Match: 14.62% Indels: 21  
DB: 29 Gaps: 1

US-09-914-698-1 (1-1861) x CNS00G5G (1-1101)

QY 1487 ValValileGlnArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyr 1506  
DB 4 GTAGTATCCAGSCCGTTCGTCTCGTGTGCTATGTAAGGCGACAGGATTAC 63  
QY 1507 GlnLeuileGlnSerValileLeuValGlnArgLysPheArgAlaAsnArgSerMet 1526  
DB 64 CAATTAATCCATCTCTGTGATCCTGGGCGAGCGCAAAATTCGCTGCTAATCGCAGCATG 123  
QY 1527 LysGlnAlaArgGlnGluPheValGlnLeuArgThrileAlaValHisLeuGlnGlnLys 1546  
DB 124 AAGCAGCAGCGCCAGGAATTTGTCAGCTCGTACTATTGCGATTCGATTTGCCAACAAAG 183  
QY 1547 PheArgGlyLysArgLeuMetileGlnArgAsnCysPheGlnLeuLeuArgCysSer 1566  
DB 184 TTCGTGGCAAGCGTCTAATGATGAGCAGGTAATTTGTTCCAACTGCTCCGCTGTTC 243  
QY 1567 MetProGlyPheGlnAlaArgAlaArgLysPheMetAlaArgLysArgPheGlnAlaLeu 1586  
DB 244 ATGCGCGGCTTCCAGGCAAGCGTCCCGCGGCTTTATGCTCGCAAAACGATTCAGGCGCTG 303  
QY 1587 MetProGluMetMetAspLeuileArgGlnLysArgAlaAlaLysValileGlnArg 1606  
DB 304 ATGACACCCGAGATGATGACCTATCCGCAAGCGCGCCGCGCAAGGTTATACAGCT 363  
QY 1607 TyrTipArgGlyTyrLeuileArgArgGlnLysHisGlnGlyLeuLeuAspilleArg 1626  
DB 364 TACTGGCGGGCTATCTAATCCGACGAGCTCAGAACGACCGAGGCTCTCTTGATATCCGG 423  
QY 1627 LysArgileAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLysVal 1646  
DB 424 AAGCGTATTGCTCAGCTGCGGCAAGGCAAGCGGTAACCTCTGTGCGCTGCAAAAGTC 483

QY 1647 GlnGluAlaValArgPheLeuArgGlyArgPheileAlaSerAspAlaLeuAlaValLeu 1666  
DB 484 CAGGAGCGGTCGCTTCTTCGCGACGCTTTATCGCATCATGCGTATACA-GTGTCA 542  
QY 1667 SerGlnLeu----- 1669  
DB 543 AGTCGATTGGTAAAGTATCCAGTACGTAGTCTTATTATAAATAATCCTCTTAATA 602  
QY 1670 -----AspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPhe 1686  
DB 603 ATCCCGCGAGATCGTCTTTTCGCGACCTGTCACACCTGCTCATGTGGTGTTCGGAGTTC 662  
QY 1687 MetSerThrPheCysTyrGlyileMetAlaGlnAlaileArgSerGluValAspLysGln 1706  
DB 663 ATGTCACAGTTTGTCTATGTCATCGATCGCTCAGCCATTCGATCAGAGGTGGATAGCAG 722  
QY 1707 LeuileGluArgCysSerArgileileLeuAsnLeuAlaArgTyrAsnSerThrVal 1726  
DB 723 CTTATAGAGCGCTGCGAGCGGATCATCTAAATTTGCGCGCTACATAATAGCACCGGTG 782  
QY 1727 AsnThrPheGlnGluGlyLeuValThrileAlaGlnMetLeuLeuArgTyrCysAsp 1746  
DB 783 AACACGTTCCAGAGGCGGCTTTGGTCACCATTTGCCAGATGTTATTCGCTGGTGGAC 842  
QY 1747 LysAspSerGluilePheAsnThrLeuCysThrLeuileTrpValPheAlaHisCysPro 1766  
DB 843 AAGACAGTGTGATATTCACACTTTGTGACCTCATTTGGGTATTCGCCCATCTGTCT 902  
QY 1767 LysLysArgLysile 1771  
DB 903 AAAAAACGAAAGGTA 917

RESULT 2

CNS00LY6

LOCUS

DEFINITION

1022 bp DNA linear GSS 14-JUN-1999

Drosophila melanogaster genome survey sequence TET3 end of BAC:

BACR48008 of RPCL-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL078898

AL078898.1 GI:5102189

GSS.

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osogawa and

Aaron Mamoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCL-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

source

1. 1022

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR48008"  
/clone.lib="RPCL-98"  
/note="end : TET3"

## ORIGIN

## Alignment Scores:

Pred. No.: 9,08e-85 Length: 1022  
Score: 1115.00 Matches: 228  
Percent Similarity: 82.6% Conservativity: 6  
Best Local Similarity: 80.5% Mismatches: 28  
Query Match: 11.72% Indels: 21  
DB: 29 Gaps: 1

US-09-914-698-1 (1-1861) x CNS00LY6 (1-1022)

QY 1512 SerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGln 1531  
Db 1 TCTGARATCCTGGAGCAGCGCAAAATWCCGTGTGTAATCGCAGCATGAAGCAGCAGCCAG 60  
QY 1532 GluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArg 1551  
Db 61 GAATGTGTCCAGCTGGTACTATTTCAGTCCATTGTCACAAACAAAATTCGTGGCAAGGT 120  
QY 1552 LeuMetIleGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGln 1571  
Db 121 CTAATGATTGARCARGTAATGTWTCACACWGCTCGCTGTTCATGCGGGCTTCCAG 180  
QY 1572 AlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMet 1591  
Db 181 GCACGTGCGCGGGCTTTATGCTCGCAAAATAATCCAGGCCCTGAWGAMACCCGARAWA 240  
QY 1592 MetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyr 1611  
Db 241 AWGACCTCAWCCGCGCARAAGCGCGCCGCAAGGTTAAACRCGTTACAGCGGGGCTAT 300  
QY 1612 LeuLeuArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLys-ArgIleAlaGln 1631  
Db 301 CTAACCCGAMGAMRTCAAAAGCACCAAGCTCTCTTGGATATCCGGAAGAGGTATTACTCA 360  
QY 1631 nLeuArgGlnGlnAlaLysAlaValAsnSerValArgCysLysValGlnGlnAlaValAr 1651  
Db 361 ACTGGGCAAAAGCAAAAGCGGTAAACWCTATGCGCTGCAAGTCCAGGAGCGGTGCG 420  
QY 1651 qPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu- 1669  
Db 421 TTTCTTTCGGGACGCTTTATCCGATCAGATCGGTASAGTTGCTTAAGTCGATTGGGTAA 480  
QY 1670 -----AspAr 1671  
Db 481 GTGATCCAGCTAGTGTAGTCTTATTATACTAATCCTCTTAATAATCCCGCAGATCG 540  
QY 1671 gLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCy 1691  
Db 541 TCTTTTCGGGCACTGTGCCACACCTCTCATGTGGTGTTCGGAGTTTCATGTCCACGTTTTK 600  
QY 1691 sTyrGlyIleMetAlaGlnAlaIleArgSerGlnValAspLysGlnLeuIleGluArgCy 1711  
Db 601 CTAIGGCATCATGGCTAGGCCATTCGATCAGAGTGGATAGCAGCTTATAGCGGCTG 660  
QY 1711 sSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGln 1731  
Db 661 CAGCGGATCATCTTAATKTBGCGCGCTACAAATAGCACCGGTAACACGTTCCASGA 720  
QY 1731 uGlyGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluI 1751  
Db 721 GGGCGGTTTTGTCAACATTTGCCAGATGTATTTCGGCTGGTGGCGCAAAAGACAGTGGAT 780  
QY 1751 ePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLysI 1771  
Db 781 ATTCAASACTTTGTGACCCCTCATTTGGKATTCGCCCACTGTCTCTAAAAAGCGAAGKA 840  
QY 1771 eIleHis 1773  
: : : : :

Db 841 GCTAAAT 847

## RESULT 3

AI519740

## LOCUS

## DEFINITION

AI519740 596 bp mRNA linear EST 02-DEC-2003  
LD39479.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
melanogaster cDNA clone LD39479 5 similar to asp: FBan0006875  
GO: microtubule binding (GO:0008017); microtubule binding  
(GO:0008017); cytoskeletal structural protein (GO:0005200);  
microtubule associated protein (GO:0005875); microtubule located on:  
3R 96A23-96A23; 08/13/2002, mRNA sequence.

## ACCESSION

AI519740

## VERSION

AI519740.1 GI:4425594

## KEYWORDS

EST.

## SOURCE

Drosophila melanogaster (fruit fly)

## ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 596)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Other ESTs: LD39479.3prime

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

hit genomic AE003749: arm:3R [20443564,20672042]

estimated-cyto:96A20-96B10: 04/10/2001

Plate: LD.394 row: G column: 7

High quality sequence stop: 453

POLYA=No.

Location/Qualifiers

1..596

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clone="LD39479"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="xfl Blue"

/clone\_lib="LD Drosophila melanogaster embryo pOT2"

/note="Organ: embryo; Vector: pOT2; Site 1: EcoRI; Site 2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2."

ORIGIN

## Alignment Scores:

Pred. No.: 1.85e-75 Length: 596

Score: 1002.00 Matches: 197

Percent Similarity: 99.49% Conservativity: 0

Best Local Similarity: 99.49% Mismatches: 1

Query Match: 10.53% Indels: 0

DB: 9 Gaps: 0

US-09-914-698-1 (1-1861) x AI519740 (1-596)

QY 607 ProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSerAspArg 626

Db 1 CCATGCTCCAAAGTGGCGGTATATGTCAACAGCAAGCTCTGCGCATCCGTAGCATCGT 60

QY 627 AsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLysCysPheAsn 646

Db 61 AATCTTCACTAGACGTAGTATTATGCAACGCCACCATCTTGAGCTGCTGCTTCTTCAAT 120

QY 647 ProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysIleGlnMetGlnSer 666

Db 121 CCCCTTTGGCTGGCTTGGACTGGAGTGTCTTTGGCGAGAGATCCAGATGCAGTCT 180



```

REFERENCE
AUTHORS      1 (bases 1 to 718)
              Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
              Lewis,S. and Rubin,G.M.
TITLE        BDBG/HIMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
              BDBG
              Lawrence Berkeley National Lab
              One Cyclotron Rd, Berkeley, CA 94720, USA
              Fax: 510 486 6798
              Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
              hit genomic AB003749; arm:3R [20443564,20672042]
              estimated-cyto:96A20-96B10: 05/19/2001
              Plate: SD.254 row: C column: 12
              High quality sequence scop: 675.
              Location/Qualifiers
                1..718
                  /organism="Drosophila melanogaster"
                  /mol_type="mRNA"
                  /db_xref="taxon:7227"
                  /clone="SD25436"
                  /lab_host="DHS-alpha"
                  /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                  culture pOT2"
                  /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
                  fractionated cDNAs were directly ligated into pOT2.
                  Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.:      2,446-62      Length:      718
Score:          851.00      Matches:    176
Percent Similarity: 85.8%      Conservative: 0
Best Local Similarity: 85.8%      Mismatches: 2
Query Match:    8.94%      Indels:    28
DB:             12      Gaps:      1

US-09-914-698-1 (1-1861) x B1642052 (1-718)

QY      1 MetGluLeuValTrpSerProValLeuGluValAlaCysGlyGluThrLeuGlnLeuIle 20
DB      614 ATGGAGCTAGTGTGGAGCNCCTCTCTTGGAGTGGCTGCAAGGAGCGCTGCAGCTAATA 555

QY      21 AspAsnArgAsnPheArgGlyGluValMetIleLeuLysSerLysSerGlnPro 40
DB      554 GACACCGCACTTCGAAAGAGGAGTGATGATCACTCAAGTCCAGAGCAACCGCG 495

QY      41 Val----- 41
DB      494 GTCAAGGTCAGTCARAAATGCAATTGTCACCCCTTTTACTTACATACCTCTTTTGCAT 435

QY      42 -----LysAsnProArgLysPheProThrValGlyLysThrIle 54
DB      434 GTATCTCTTACCTTTCTACCTGAAGAACCGCGCAAAATTTCTACTGTGCGCAAGACCT 375

QY      54 uGlnLeuLysSerProThrGlyValGlyLysThrMetLysSerValValSerAlaAlaVa 74
DB      374 CGAGCTGAATGCCACAGGAGCTGGCAAGCAATGAAGAGCGTGATTCGCTCTGT 315

QY      74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94
DB      314 GCAGCAAGAGAGCGATGCTCTGCAGCAGCAGCGCTCCCTCCACGACATGCGGAGT 255

QY      94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValG1 114
DB      254 GACTGCTCTTCCGTCGCGTCGATGGGCACATCCACCTCCACAGGCTCCTCTTGTGCA 195

QY      114 uLysAsnValTyrLysThrProGlnGluProValTyrLysSerProGlnProArgSe 134
DB      194 GAAGAATGTATACAAGACTCCACAAAGAGCCGCTATACATATCACCAGGCTCGCAG 135

QY      134 rLeuLysGluAsnLysSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154

```

```

Db      134 TCTTAAGAAATCTAAGCCCTATGACGCCAGAAACCTCGAGTGATTGACAACT 75
QY      154 uArgPheThrProLeuThrGluThrArgGlyGlyGlnAlaThrIlePheProAspAs 174
Db      74 GCGATTACACCTCTCCAGAAACCGG-GGCAAGAGCAAGCTACATTTTCCCGACA 16
QY      174 nLeuAlaAlaTrp 178
Db      15 TCTGGCAGCCTGG 3

RESULT 6
BC026699      3130 bp      mRNA      linear      HTC 20-SEP-2002
LOCUS      Mus musculus, clone IMAGE:3994865, mRNA.
DEFINITION      BC026699
ACCESSION      BC026699.1 GI:19934325
VERSION      HTc.
KEYWORDS      Mus musculus (house mouse)
ORGANISM      Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 3130)
AUTHORS      Strausberg,R.
TITLE      Direct Submission
JOURNAL      Submitted (04-APR-2002) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
              Email: cgabbs-remail.nih.gov
              Tissue procurement: Gilbert Smith, Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Baylor College of Medicine Human Genome
              Sequencing Center
              Center code: BCM-HGSC
              Web site: http://www.hgsc.bcm.tmc.edu/cdna/
              Contact: amg@bcm.tmc.edu
              Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
              Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
              Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 40 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: no polyA-tail.

FEATURES
source      1..3130
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="CZECH II"
              /db_xref="taxon:10090"
              /clone="IMAGE:3994865"
              /tissue_type="Mammary tumor metastasized to lung. Tumor
              arose spontaneously from a senescent normal mammary
              (clonal) outgrowth infected with the virus MMV."
              /clone_lib="NCI CGAP_Lu29"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.:      2,826-50      Length:      3130
Score:          723.50      Matches:    235
Percent Similarity: 43.69%      Conservative: 177
Best Local Similarity: 24.92%      Mismatches: 270
Query Match:    7.60%      Indels:    261
DB:             11      Gaps:      27

```

US-09-914-698-1 (1-1961) x BC026699 (1-3130)

Qy	878	LysPheHisAlaAlaThrValLeuGln-----LysTrpTrpArgArgHisTyr	894	Qy	1208	-----AsnTyrLeuHisLeuGlnThrThrThrLysArg	1211
Db	369	AAGCTGAATCCTCATCCCTTGAATCCAGTTTATGTTTCAAGAGGTGGAAGCGCAAG	428	Db	1374	GCTGCCATTCTCTACAGTCTTATTTTCAGAAATGAGAACAGCGGCGAGCTACCTGAAG	1433
Qy	895	LeuHis-----ValValleGlnArgArgileArgHisLysGluLeu	908	Qy	1219	-----IleGlnIleLysPheArgAlaLysArgGluMetLysLys	1231
Db	429	CTGCAGTTACAAACCAAGCTGCAGTACCATTCGACGAGCTTTTACAGAAATGCGCATCTC	488	Db	1434	ATGTGCAAGCGCGCTCTTGTTCATTTTCAGAGTTTCTACTGTGCTTACAGAGCACAAATCAGT	1493
Qy	909	MetArg-----ArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHis---	925	Qy	1232	GlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValValGln-----	1247
Db	489	AGGAAACAAATACGAGAAAGATCGCTGTGTGATACAGTCGTGTACAGATCCACAGG	548	Db	1494	CAGAAGAAAGACTTCTTGTGAGCTCAAGGAGCAGCAGCTATTTTGTTTGAGCAGCTTACAGA	1553
Qy	926	GlnMetArgTyrTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeu	945	Qy	1248	-----LysArgArgArgAlaLeuLeuGlnMet-----	1256
Db	549	GAATTGCAGAAATACATTCATATT-----AGGCTTGTGTTATGTGCATC	593	Db	1554	GGCTGCAAAAGTACGCGGCAAAATCAAGCAGCAGCTTACGGCTGCTGTTCAAAATTCAGAGG	1613
Qy	946	GlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle	965	Qy	1257	-----ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLys	1272
Db	594	CAGAGAGAGTTCGGTCTTTCAAGCCCAAAAGTTATATAGAGAAAGAGATGCCATC	653	Db	1614	GTTTTTAGAGCCACAGCTCAGAGGATGAATACCACTATGCTTCAGTCGCTGTGAAG	1673
Qy	966	IleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPhe	985	Qy	1273	LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly	1292
Db	654	CTGACTCTCAGAGACATTACAGGCGCAGTCAGAAAGGGAAGTAGCACACGCTGACTAT	713	Db	1674	ATTTCAGAGATGTACAGGCGCACACAGAGTTGCTTATGACATGAGAATTCAATTTCTAAAG	1733
Qy	986	ValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAla	1005	Qy	1293	ThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsn-----	1306
Db	714	TTGCAGAAACGAGCTGCCACCATTCGACTCAGCTCGCTTCAGGGGG-----	761	Db	1734	ACTAGAGAGCTGTGTTTGCCTCCAGTCTGCCCTACCGTGGCTGGCAGGTTCCGCGACAG	1793
Qy	1006	LysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnAla	1025	Qy	1306	-----HisLeuLeuArg-----	1310
Db	762	-----ATGAAGGCTGCCATTCATACAGGTTG-----CAGATTGGA	797	Db	1794	CTCAGGAGGCAACATGAGGCTGCAGTGAAGATCCAGTCTACGTTTAGATGGCCGTGGCC	1853
Qy	1026	AlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHis	1045	Qy	1307	-----HisLeuLeuArg-----	1310
Db	798	GCTGCTGTGTTCTTCAGTCTTATGGAATAGAGACAG-----GAAAGAGTT	845	Db	1854	CAGCAACAGTATAAATCTACTAAAGCAGCTGCAGCAGCTCATCCAGCAGAGGTGAGGCA	1913
Qy	1046	GluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTyr	1065	Qy	1311	-----LysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThr	1326
Db	846	AGATTTTAAACCTTAAAGATGGTTATTAAAGTTCGAAGCACATAGAAAA-----	899	Db	1914	CGGGCTGACGGAAGGAGGAGCAGCTTGGGTACATCCAGTCCGTATGCTCAGCTCTGGTG	1973
Qy	1066	SerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleGln	1085	Qy	1327	LeuGlnArgArgTyrArgAlaArgLeu-----	1335
Db	900	-----TACCAACAACCTCAAGAAATATAGAAATAAGAAAGCAGCCATTACAATTCA	953	Db	1974	TTTCAGGCTGCATGGAAGGAAAGATGCTGAGAAGACAGATTGCCCGCAGCATCAATGC	2033
Qy	1086	GlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArg	1105	Qy	1336	-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu	1347
Db	954	ACTCATTTCCGAGCTTCCATTTCAGCCAGAGAGTCTAGCATCTTATCAGAAAAACAGG	1013	Db	2034	GCTGCCTCATACAGTCATATTACAGGATGCATTCAGCGCAGAGGTGGAGCATCATG	2093
Qy	1106	SerSerValLeuLysValGln-----AlaTyrArg	1115	Qy	1348	LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLeuAlaMetGlnLysGln	1367
Db	1014	TCTTCAGTCATTCTTACAGTCTGCATGTAGAGGATGCAAGCCAGAAAGCATTTCCG	1073	Db	2094	AAAAACGCTGCCCTTCAGATTTCAGCTGTGCTATAGAGCTTACAAAGTTGGAAGAACAG	2153
Qy	1116	LysAlaThrIleGlnMetArgIleAspArgAsnHisTyr-----TyrSerLeuArgLys	1133	Qy	1368	ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArgGly	1387
Db	1074	CATGCTTTAGATCCCTGTATATAAGATTCAAGTATATATGCGGCTTATATATAGGAAG	1133	Db	2154	CGCCACTTATCTTGAACAAACAAAGCAGCTGTAGTAACCTTTACAGTCAGCTTACCGTGT	2213
Qy	1134	-----AsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGlu	1150	Qy	1388	AsnLeuGluMetArgLysGlnIle-----	1395
Db	1134	ACATTTTCAGAACTTCAAAAATGCCAATAAAGCTGCAGTCAATTGTCAAAATGAACAA	1193	Db	2214	-----ATGAAAGTGAAGACGGGTGCTGATGCCAACAGCAGCGTAATATACAGTCT	2270
Qy	1151	GlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArg	1170	Qy	1396	-----GluValTyrGlnLysGlnArgGlnAlaValIleArg	1407
Db	1194	TCACGAAGCAGGATTTACAAATAAGACAGCTGCTGCTGTTATCCAGCGGTGGTACCGT	1253	Db	2271	AAATTTAGAGCCTACAGAACCAAAAGAAAGTACACGACTATAGAACTTCAGCTATTGTA	2330
Qy	1171	MetArgGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle	1190	Qy	1408	LeuGlnLysTyrTrpArgSerIleArg-----	1416
Db	1254	TCCCAAAAACCTGGCCTCACAAAGAGAAAGAGTACATACAGTGGGTGATCTGTATC	1313	Db	2331	ATTTCAGAGTGTACCGGAATATTAATTTACACCCAGCAACATCAGGAATACT	

QY 1417 -----AspMetArgLeuCys 1421  
Db 2451 ATTGACACATGACATGCGGCGCACACTTATTAAAGCCATGTTTAAAGATGGTCAGTCT 2510  
QY 1422 LysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAla 1441  
Db 2511 AGAGTAAGGTACTCTCAAGATGAGACAGCAGCTCTTATTAATTCAGGTGAGATACAGCG 2570  
QY 1442 ThrValGlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeu 1461  
Db 2571 TATTATCTAGTAAAGTTCAGATGAGAAAGTCTTGAAGGCTATTAAACC 2630  
QY 1462 MetGlnAlaPheIleArg----- 1467  
Db 2631 CTTCAAGCTGGAGTCAGAGGACCAAGAGTGGAGCGGACGTGTAAGAAAGTGCATTTCGCA 2690  
QY 1468 AlaThrLeuLeuMet-----ArgGlnGlnArgArgGluPheGluMetLys 1482  
Db 2691 GCACACACTCATTCAGTCACATTTTATAGGACACAGCGCAGCAAACTTACTTTCATCGGTG 2750  
QY 1483 ArgArgAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAla 1502  
Db 2751 AGAAGCGGCGCAACCATGTCAGACAGATACCGAGCAGTGAAGAGGAGCGCTGAG 2810  
QY 1503 ArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAla 1522  
Db 2811 TTTCAGAGGTACAGCAGACTAAGCGCTTCTGTACTCTCATCCAGGCTGCCTTCAGGGA 2870  
QY 1523 AsnArgSerMetLysGln----- 1528  
Db 2871 CTGAGAACTAGGAGACACTTAAAGCTATGCACCTGGCTGCTCTCTCATTCAGAGGAGA 2930  
QY 1529 -----AlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 2931 TTCAGGACTTTCGCATGAGAGAAAGTTCTTTCCTCCAGAAACTGCCATTTGGATT 2990  
QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db 2991 CAGAGCAGTACCGTGCA---AGACTTATGCGCAAGTATTCGCGACACAGTACTGTG 3047  
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 3048 GAAAGAGGAGGTATTAAGATCCAGTCATCATCAGAGGATGGTGGTAAAGGAGGAGT 3107  
QY 1584 GlnAlaLeu 1586  
Db 3108 CAAAGATG 3116  
RESULT 7  
AK052326  
LOCUS  
DEFINITION  
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D330028K02 product:hypothetical transglutaminase/calponin-homology domain, CH-domain structure containing protein, full insert sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK052326.1 GI:26095125  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
MEDLINE  
99279253  
PUBMED  
10349636  
REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3983)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers  
1. .3983  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:D330028K02"  
/db\_xref="MGI:2421386"  
/db\_xref="taxon:10090"  
/clone="D330028K02"  
/tissue type="heart"  
/clone lib="RIKEN full-length enriched mouse cDNA library"  
dev stage="13 days embryo"  
1. .3983  
/note="hypothetical Transglutaminase/Calponin-homology domain, CH-domain structure containing protein (SCOP|49309, SCOP|47576, evidence: SCOP)"  
misc\_feature  
ORIGIN

Alignment Scores:		1-71e-49	Length:	3983
Pred. No.:	716.50	Matches:	268	
Percent Similarity:	43.11%	Conservative:	167	
Best Local Similarity:	26.56%	Mismatches:	388	
Query Match:	7.53%	Indels:	186	
DB:	11	Gaps:	39	
US-09-914-698-1 (1-1861) x AK052326 (1-3983)				
QY	1	MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuLeuLeu	20	
DB	452	ATTTCCGTTACCTGCGACCGCACTGCGAGAGGGGAGTGGAGGAGATTGTTCATATTCTT	511	
QY	21	AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro	40	
DB	512	GTAAAT---GATTCTCTGAAGCCACGAGTATATTA---GGAAATGCGAAGAGCCT	565	
QY	41	ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr	60	
DB	566	AAGAAGAAAGAGAGTCTTTTGGATACCAGTAAG-----AAGATTCCAGCC	613	
QY	61	GlyAlaGlyLysThrMetLysSerVal-----ValSerAlaAlaVal	74	
DB	614	TCTTCAAAACATACAAAAGAGACTTCCAAAACCAACATTTTAATGAATCATTTACTATT	673	
QY	75	GlnGlnLysLysArg-----MetSerAla	82	
DB	674	TCACAAAGACAGAAATTAGAGCCCACTGACGCTTGTGAAATCTGGCTATGAGTAA	733	
QY	83	AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSer---ArgProAla	101	
DB	734	TGCTCTTCCCAACAGAAAC-----AAAGTCCCAACCCATCATTAGTCTCTATT	784	
QY	102	AlaTrpAlaHisProProGlnAlaProLeuVal-----GluLysAsnValTyrLys	119	
DB	785	AGAGAAATGCCAGAGTGAACCTGCTGCCACTGTTTACGCGAGTCCACTGCCTATTCA	844	
QY	120	ThrProGlnGluProValTyrIleSerProGlnProArgSerLeuLys---GluAsn	138	
DB	845	TCTTCTTCATGAA-----TCTGAAATAACACAAATTTTAAAGATACAGAT	889	
QY	139	LeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPhe---Thr	157	
DB	890	GCCAGCATTTCAAACTTTTGAATTTAATGAGGAAGTCGCAATGAACTTTTATTATAT	949	
QY	158	ProLeuThr-----GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsn	174	
DB	950	CCCATTAGTGTCTGTCCACAGAGTGAAGGGGATAGGAACTCAGCTTGCCTCCCAAC-	1006	
QY	175	LeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAsp---	193	
DB	1007	---TGTTCTTCACCTTTGAATAGTACACAGCCCAATACATCTTCTAAGTCCGATCT	1063	
QY	194	---MetArgProArgGlyLeuThrProAspAspLeuGluAsp---GlnProAlaThrAsn	211	
DB	1064	TTTGTAAACACAGGTATACGTCTGATAATGACCTGAAATCAATGAAATGTTTTATCA	1123	
QY	212	LysThrPheAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp---	230	
DB	1124	GACACATTTAGGAAGATCCCGCAGAGTCTGTATGTCTGGAATCCCAACTGTCCATGAA	1183	
QY	231	---CysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal	249	
DB	1184	GTTTGCCAAACATC---TTAAGCCAGATCTTTTTTAATGATATATGACTTAA	1240	
QY	250	HisAlaThrHisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerPro	269	
DB	1241	AAAGGGCTAAACTCAAGTCGGTC-----AATCCCGTT	1273	
QY	270	ArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySer	289	
DB	1274	TTGTCGCCAACACAAATTT-----GTAAAGGATAGCATGGGACACAGTAGGTACG	1321	
QY	290	ProLeuArgLysTyrSerGluSerMetLysAspLeuSerLeu-----Leu	304	
DB	1322	CAACCGGTAAAGTCAATGAAGCTTCTCAGGATTGGAGATAAATGAAGGTTGGCGTAT	1381	
QY	305	SerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIle	324	
DB	1382	ACCCCGAGTGTGAGCATGCA-----CAGACTCCAGCAGCAGATCTGAAAAACAA	1432	
QY	325	ArgSerIleGlu-----	328	
DB	1433	AACCTGTGAAGTAGAAGCAGACATACGTATGATTTTCAAGCAAAAGCCCAAGATTCT	1492	
QY	329	-----GlnAsnArgTyrTyrGlnGluGlnGlnIleGlnIleLysAlaLysAspLeuAsn	346	
DB	1493	GAGTTTCAGATGCTTTTGTCCACAGAGTAACACCCGACACAGAGACCCCAATTCTT	1552	
QY	347	SerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSer	366	
DB	1553	TCTGCCACTGTACGAAAGAAACCCACCAATGCCAGAGAAAGCTA-----CCT	1603	
QY	367	GluIleLeuAlaGlnSerSerArgPheAsnLeuHis-----	378	
DB	1604	GAGATTAAACAGCCAGACGCAAAAGATGTCTCAAGGTCTAGTAGTGACCTGGAAAA	1663	
QY	379	GluValGlyArgLysSerValLysGly-----SerProValLysAsnPro---	393	
DB	1664	GAGTGGTAGCTGAGAGAGAAAGGTTTTCATTCTCTTCCCGTTGGAGCCAGGA	1723	
QY	394	-----His	394	
DB	1724	GTGAGTAAAGCTCTGAGCTACAGGATGAAGTAACGCTGCAACTGTTGCTGCTCGG	1783	
QY	395	LysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArg	414	
DB	1784	AAAGAAAGAGCCACCGA---ACATGGGAGATGCA-----AATGGAAGTAGCAGCT	1834	
QY	415	AsnGluThrValAlaIleSerProLysLys-----GlnArgVal	428	
DB	1835	GAAGATGGATGGACATGTGTGAAGTCAACGAATCCATTTTCTCTCTGGAATCTACA	1894	
QY	429	GluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSer	448	
DB	1895	CCCTCAACTGTTCAGAACACCAAGAAAGAGGGCATACCTCAAAGCGGATAAGCAGC	1954	
QY	449	AlaHisAlaTrpProHisAlaGlnSerLysPheLysLeuAlaGlnThrMetSerLeu	468	
DB	1955	-----CTGGAGCGGTGAGGCTGAGAGAAATGGATTTCATCAATA	1996	
QY	469	MetLysLysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLys	488	
DB	1997	TTAAACACCTCTTTCTAAGACAAAGAAAGAGAGATCCATT-----GTTGCT	2047	
QY	489	LeuTyrAspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThr	508	
DB	2048	GTAGCACAGTCTCATTTG-----ACCTTTATAAGCCATTAAAGCAGCTATTCTCT	2098	
QY	509	ThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArg	528	
DB	2099	AGACCCCAATGCGTTTGTCTCAAGAACACATGTTCTATGACGACGCTGGAAGGAAAG	2158	
QY	529	HisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAsp	548	
DB	2159	CAGGAACAGGGCTTCACTTGGTGGCTAAATATATATTAACA---CCTGATGATCTTCACT	2215	
QY	549	AlaAspLeuAsn---AsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLys	567	
DB	2216	GTAAACCAATGTTCTTAAAGTGAACGCTGCCAGCCTGTTTGGAGCGGGAAGCCAA	2275	
QY	568	GluLeuVal-----ValAlaProThrLysGluGlnSerMetAsn---TyrLeu	583	
DB	2276	CATAAATAAGTGTCCCGAAAGACCTACTTAAGAGGAAGTGTCTCTCAGAGCTTATACA	2335	



Qy 584 ThrLysTyrArgLeuGluThrLeuArgLysAlaValGluLeuPhePheSerGluGln 603  
 Db 2336 GCCAGCTCGAGCTGAATAGATTACCGGTACAGCTTGCTCTGTTTACTTCTGAAATA 2395  
 Qy 604 MetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgLysLeu 623  
 Db 2396 ATGGTTAAAGCTATTAAAAAGGTTGAAATCGAAATGAAAGTTGGCGGCTACTTGTCTGA 2455  
 Qy 624 SerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeu 643  
 Db 2456 AAAGACCCACACCTCTGGAAGATATGGCAACGCCCAAAAGTCTGAATTGGCTCTGTG 2515  
 Qy 644 CysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysIleGln 663  
 Db 2516 TCATATATCTCTGTGGCTTCGAATTGGTCTGGAGACAGTTTTGGAGAACTGATCCCT 2575  
 Qy 664 MetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArg 683  
 Db 2576 CTGGCAGACACAGTGACGTACCGGATTGGCGATGTTCAATCTGAATCGCTTACTTTGG 2635  
 Qy 684 AsnLys-----CysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu--- 700  
 Db 2636 AATCTGACATAGCAGCTGAATATAGACACCCACTGTTCTCTCTCTGTTTGGGATGGT 2695  
 Qy 701 TyrAlaGluThrIleLysIleHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720  
 Db 2696 CATGAGCAGCGCTTCTTAAGTTCATATGAAATAATGCTGTTGTTGATCTGTTCTCT 2755  
 Qy 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
 Db 2756 GATCATGCTAAATTTCCAGACTTATTCATGATGATCTTGCTCTTCTGTAAGATGCT 2815  
 Qy 741 ProHisLysGluThrLysAspIleLeuArgPheSerSerGluLeuLeuAlaHisnle 760  
 Db 2816 GAATTCAGGCTAGCAAGAGCTCTTCTAGCTTTTCCAGGATTTCTCAGTGGTGAA 2875  
 Qy 761 GlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
 Db 2876 GGTGACCTTCTGTCACCTTAGCTTCTGGGTACCTGTGAGCAGCAGTTCAGACACCA 2935  
 Qy 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
 Db 2936 CTTGATGAATTTGATTTGCTGTATACAAATCTTGCAGTGGACTTGCAGTGGGTTCGC 2995  
 Qy 801 LeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
 Db 2996 CTTGTGCAACTGTGGAGCTTCTCACAGAACTGGAATCTGTCCGCAAAACTCAGATC 3055  
 Qy 821 ProAlaIleSerArgGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
 Db 3056 CTTGCAATACTCGAGTTCAGAGATGACAAATGTTGATCTGCTCTGCTTCAAGTTCAG 3115  
 Qy 841 Glu-AlaAsnPheGln----- 845  
 Db 3116 TCTCGGGAGTTCCACTGACTGACGAGCATGTGTAACGAGCGTCAGAAATAAGCTTCGTG 3175  
 Qy 846 ----- 847  
 Db 3176 TCCTTGTGAAGCTTGAATGTGTGACGTACCGTCTCTAAACGAGCTTCCCTTTCATCTGTAGG 3235  
 Qy 847 yGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLe 867  
 Db 3236 GAGCGCTATCTGCTTAGAGATGTTGTGGACGACGACAGAGAGAAACCCCTTGGTTCCT 3295  
 Qy 867 uTrpGlnLeuIleTyrLysPheArg 875  
 Db 3296 TTGGAAATAAGCACGTGGCTTTCCAG 3320  
 RESULT 8  
 BG636223  
 LOCUS  
 DEFINITION SD13838.5prime SD Drosophila melanogaster Schneider L2 cell culture

pot2 Drosophila melanogaster cDNA clone SD13838 5 similar to asp:  
 FBan0006875 'cytoskeletal structural protein' located on: 3R  
 96223-96A23:: 04/13/2001, mRNA sequence.  
 BG636223.1 GI:13763760  
 EST.  
 BG636223.1  
 ORGANISM  
 Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 589)  
 HARVEY,D., BROKSTEIN,P., HONG,L., EVANS-HOLM,M., SU,C., TSANG,G.,  
 LEWIS,S. and RUBIN,G.M.  
 BDGP/HMMI Drosophila EST Project  
 Unpublished (2001)  
 CONTACT: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd., Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
 Based upon one or more reads of this clone where vector sequence  
 was present at both ends, this clone has been determined to contain  
 contain a cDNA insert on the order of 600-1000 bases. hit genomic  
 AE003749: arm:3R [20443564,20672042] estimated-cyto:96A20-96B10:  
 04/13/2001  
 Plate: SD.138 row: D column: 2  
 High quality sequence stop: 547.  
 Location/Qualifiers  
 1..589  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="SD13838"  
 /lab\_host="DHS-alpha"  
 /clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
 culture pot2"  
 /note="vector: pot2; site 1: EcoRI; site 2: XhoI; Sized  
 fractionated cDNAs were directly ligated into pot2.  
 Plasmid cDNA library."

## FEATURES

source

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,96e-45 Length: 589  
 Score: 647.00 Matches: 119  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.80% Indels: 0  
 DB: .12 Gaps: 0

US-09-914-698-1 (1-1861) x BG636223 (1-589)

Qy 1743 ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPhe 1762  
 Db 1 CGCTGGTGGCAACAAAGACAGTGGATATTCACACTTTGTGCACCTCATTTGGTATTC 60  
 Qy 1763 AlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIle 1782  
 Db 61 GCCCACTGCTCTAAAGCGAAAGATCATTCACGACTATATGACCAACCCAGAGCCATT 120  
 Qy 1783 TyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAla 1802  
 Db 121 TACATGGTGGCGGAAACTAAGAGCTGTGGCCCGCAGGAAAGATGAAGCAAAATGCA 180  
 Qy 1803 ArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrPro 1822  
 Db 181 CGCAAGCGCGCCCAATGACAAAGTGGCGGTATTAAGACCAAGATATAAATCTTCAGCGCG 240  
 Qy 1823 CysSerLeuProSerLeuGluProAspPheGlyIleIleArgTyrSerProTyrThrPhe 1842  
 Db 241 TGTTCCTGCGCAGCTGGAGCGGAGCTTCGGAATCATCCGCTACAGTCCCTCACAGTTT 300



Db 277 TTACAGAAATGCACAGGCTGCAGCTCTCATCCAGGCTACTTTTAGA----- 324  
QY 1094 MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAla 1113  
Db 325 -----ATGCACAGAACTTACGTTGCATTCACATT 354  
QY 1114 TyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLys 1133  
Db 355 TGGAAATGTCCTCGATTAGA----- 375  
QY 1134 AsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGlu 1153  
Db 376 -----ATTACAGCAGTGTATCGAAATATACAGAACTATAAAATTACAAAAAGAA 423  
QY 1154 AsnTyrLeuArg-----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMet 1171  
Db 424 AAACCTTATCAGAGAGAAACACACAGTGTCTGTCTTAATACAAAGCAGCGTACAGGATG 483  
QY 1172 ArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsn 1191  
Db 484 TACAGGCAG-----CGTTGCTTCTATCAA 507  
QY 1192 ValGlnArgArgTyrArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHis 1211  
Db 508 -----CAGCGTCTGGTGGCAGCA----- 525  
QY 1212 LeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLys 1231  
Db 526 -----AAGTATACAAAAAATACAGAGCA-----AATAAA 558  
QY 1232 GlnArgAlaGluPheLeuGlnLeu-----LysLysValThrLeuValValGlnLysArg 1250  
Db 559 AGACGCGAAGATCTCTCTCTGTGTGTAAAGGAAGAACACCTCTTCTTCAGATGCAATTT 618  
QY 1251 ArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThr 1270  
Db 619 CAGGGTTTAAACACTGCCAAAACAGTCGCGCAGCAGCAT-----GGGGCGCGCT 666  
QY 1271 IleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyr 1290  
Db 667 ATGATTACAGAAGCATTATAGACCTTTAAAGCAAGAGGCTCATGGAGCCGAG--- 723  
QY 1291 ArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArg 1310  
Db 724 AGAGCTTCCAGCTCGA-----TGTAGA 747  
QY 1311 LysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaIleThrLeuGln----- 1328  
Db 748 AAATATAAAGCTAAGAAATACTTATCCAAAGTGAAGCTGCTGCAGGATTCAAGCTTG 807  
QY 1329 --ArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
Db 808 TATAGCGCTTGAGAGCGGAC-----AAGAAATATCTAATCTTA 846  
QY 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
Db 847 TTGAAAGCTGTTAACTATTGAAGTTACCTCTGCACAGCTGCGCAGAGGAGG--- 903  
QY 1368 ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGly 1387  
Db 904 -----TTCTTGAAGATGAGGCGCGCAGCTATTATAAATCAAAAGAAATGGAGAGCT 954  
QY 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
Db 955 ACACCTTCTGTAAGAGAGCTCGTGAGAAATTTAAAGACATCGAGCTGCTGTGTG--- 1011  
QY 1408 LeuGlnLysTyrTyrArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArgArg 1427  
Db 1012 ATTCAGCCCACTTTAGGGCTATCAG-----GCAGGAGCTCTCTCTGCGAG 1059  
QY 1428 IleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArgArg 1447  
Db 1060 CAGAGATCTGCTGTTTAAATCATACAGACATGTACGAGCCATGCTGCGAAGCAG 1119

QY 1448 GlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArg 1467  
Db 1120 GAAAGATAAGTATATCAAAATTGAAAGTCTACGGTTGTGTACAGCCCTTGGCGGT 1179  
QY 1468 AlaThrLeuLeuMetArgGlnGlnArgGluPheGluMetLysArgArg----- 1484  
Db 1180 GGC---TGGTTAGTAGGAAGGGTTTCAGAGCAGAAAGCCAAACTCGGCTTTTCCAC 1236  
QY 1485 -----AlaAlaValValIleGlnArgArgPheArgAla 1495  
Db 1237 TTTACAGCAGCTGCCCTACTGTGCATGTCTCTTAAATCCAACGAGCTTATCGGCTC 1296  
QY 1496 ArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeu 1515  
Db 1297 CAGGTGACCTTGAGGAACGCTAAGAAGCAC-----ATGGATTCACTCATCTTC 1344  
QY 1516 ValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArg-GlnGluPheValGln 1535  
Db 1345 ATTCAGAGATGCTTTTCGAAAAAAGATTACAGGAAGAGGTTTCATTGAACAGTACCACAA 1404  
QY 1535 nLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGln 1555  
Db 1405 GATC-----CTAAGCACTCGACGTGAAGCTCATGCTGTGGTCTCA 1446  
QY 1555 uGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAla 1575  
Db 1447 CAGGACAGCGCTGCATCCGTGATA-----CAGAAGCGGTGACG 1485  
QY 1575 gGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIle 1595  
Db 1486 CGCTTTCTCTCTCGCAGAGACAGAAAAAGATCACTAGC----- 1525  
QY 1595 eArgGlnLysArgAlaAlaLysValIleGlnArgTyrTyrArgGlyTyrLeuIleArgAla 1615  
Db 1526 -----TGCGCCACTTAGAAATTCAGGCATTATCGAGAGCTATTCTTGGAGGAA 1572  
QY 1615 gArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGln 1635  
Db 1573 GAAAAATGACCAACACAGAAATATCCGAGAGCTATCCGAGAGCCTTAGAGCTGTGAGCAGC 1632  
QY 1635 uAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGln 1655  
Db 1633 CGTGAGAGAGAGAACAAAGCTGTACAGAGAACTGAGCGAGCGCTTCATCCTGCTCAC 1692  
QY 1655 YargPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgTh 1675  
Db 1693 CTACAAACACTTGTGCGCCATTCTGGATCGCTTGAAGCAGCCTGGAGGTCGTTTACCAGACT 1752  
QY 1675 rValProHisLeuLeuMetTyrCysSerGluPheMetSer-----ThrPh 1690  
Db 1753 CTCTCCA-----CTTGTGTGGAATATGCTGAAGCGGAGCGGTATCTAC 1800  
QY 1690 eCystyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluAr 1710  
Db 1801 CATCTTGTGTGATCCGAAGCTGAATCGAGTGTTCTTGTATGGAAGTCGTTGGATA 1860  
QY 1710 gCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
Db 1861 TGCTGTTCAAGTATTGCTGAATGTGGCAAGATATGATAAGACGATCGCGCTGTGTACGA 1920  
QY 1730 n---GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
Db 1921 GGCCGAGAACTGCGGGACACACTGCTGGAGTCTTCGAGGTGTAACGAGAGAACCTGG 1980  
QY 1743 ---ArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyrVa 1761  
Db 1981 AGACCGAGTCTCGAGAGAGTGCAGCATTTTACAGAAGACTTGTCTTGTGTAGCAGT 2040  
QY 1761 lPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAl 1781  
Db 2041 GCTGTGTG-----AAGACAGAGCAGTGTGCTTTTGTATGCACAAAGTAGATCTAAAGT 2091

```

QY 1781 alleTyMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAs 1801
Db 2092 CACTGATCGATTATTCGCTCTACAAATTTACAGTTCCTTAAGCATAAAGTAAT-----2146
QY 1801 nAlaArgLysProProMetThrSerGlyArgTyTrLysSerGlnLysIleAsnPheTh 1821
Db 2147 -----ACACAGGATTTGTTTGATAGCAGAGAGCAATTCCTG 2184

QY 1821 r-----ProCysSerLeuProSer-----Ile 1828
Db 2185 TGTAGGCTTTCTTGT-----ATACCAAGAAAGAACTATGAAGCACCAGGTAGTTTCAAGACT 2241
QY 1828 uGluProAspPheGlyIleArgTyTrSerProTyTrThrPheIleSerValTyAl 1848
Db 2242 TAGCCACAGTGGTGTGGGGCTGTATATGTGGAGAAATACAAACTCCCTGCAAGC 2301
QY 1848 aPheAspThrIleLeuCysLysLeuGlnIle 1858
Db 2302 TATTCAACTGGTAATGATACACTTGGAAAT 2332

RESULT 10
BH374743/c
LOCUS BH374743 485 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-143H23.TF NP-TAM Anopheles gambiae genomic clone
AG-ND-143H23, genomic survey sequence.
ACCESSION BH374743.1 GI:17320885
VERSION BH374743
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 485)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
12655398
Other GSSs: AG-ND-143H23.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(IGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..485
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-143H23"
/clone_lib="ND-TAM"
/note="Vector: pBCBAC1; Site_1: HindIII"

ORIGIN

```

```

Alignment Scores: 2.34e-26 Length: 485
Pred. No.: 429.00 Matches: 78
Score: 71.25% Conservative: 36
Percent Similarity: 48.75% Mismatches: 46
Best Local Similarity: 4.51% Indels: 0
Query Match: 28 Gaps: 0
DB:

US-09-914-698-1 (1-1861) x BH374743 (1-485)
QY 848 GlyAspIlealaalaGlnaspIleValAspGlyHisArgGluLysThrLeuSerLeuLeu 867
Db 482 GGCACATCCCGCCCAAGGACATCTCGACGGGCATCGGAGCAGACGATGTCCTGCTC 423
QY 868 TrpGlnLeuIleTyTrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln 887
Db 422 TGCACATTTGTACAAATTCGTCGGCCAAATTCATCGCGCCGCAATCGTGTGCAG 363
QY 888 LysTrpTrpArgArgHisTrpLeuHisValIleGlnArgArgIleArgHisLysGlu 907
Db 362 CGCTGGTGGCTATGAACCTGGCTGAAGTGACCATTTCCCGCGGATCGAAGAAACGCT 303
QY 908 LeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMet 927
Db 302 GCCTTGGCGCGAGAACGGCGGCCCGCACGATTCAGGCCGCTCTCGCGGGTACTGCGTG 243
QY 928 ArgLysTyTrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947
Db 242 CGCGTTTGGTACGAAGCCCATCGACGGCAGAACGTGGCGCTATCGACGATCAGCGG 183
QY 948 PheThrArgArgTyTrLeuAlaGlnLysGlnLeuTyTrGlnSerTyTrHisSerIleThr 967
Db 182 TTCTCCAGACGATATCTGGCGCAGAAAGTGGTGGTCGCCGCGGTTCAGCGCATCGTCGC 123
QY 968 IleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGlu 987
Db 122 ATACAGCAATGTTGGCGAACGGTGGCCAAATCGCCAGCGCGGGAGCGTTTCTGCTC 63
QY 988 LeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLys 1007
Db 62 TGCAGAAAGTCACAAATCTGACTGCAAAACATCTACGACGGTACGGCTCGGCAGGAAG 3

RESULT 11
CNS04QCR 1008 bp DNA linear GSS 01-SEP-2000
LOCUS CNS04QCR/c
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
129E12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL302436
VERSION AL302436.1 GI:8181338
KEYWORDS GSS; Genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
JOURNAL

```

MEDLINE 20359837  
 PUBMED 10899143  
 REFERENCE 3 (bases 1 to 1008)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
 - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.  
 FEATURES  
 Location/Qualifiers  
 1..1008  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone\_lib="129E12"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BG129BC06SP1-end :  
 PUC-Ori"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,77e-22 Length: 1008  
 Score: 388.00 Matches: 125  
 Percent Similarity: 46.46% Conservative: 72  
 Best Local Similarity: 29.48% Mismatches: 126  
 Query Match: 4.08% Indels: 101  
 DB: 29 Gaps: 14  
 US-09-914-698-1 (1-1861) x CNS04QCR (1-1008)  
 QY 1133 LysAsnValIleCysLeu-GlnGluArgLeuArgAlaIleMetLysMetArgGluGlnAr 1152  
 DB 996 AAACCTGCTTTTGTGTCACACAGAGGTTCAGACGCTGACGCTGATGAGAAAGCTCG 937  
 QY 1152 gGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetAr 1172  
 DB 936 TTCAGACTACTGTCTACAGCGCAGGCGAGTATTTCCCTTCAGGAGCCTCGAGA----- 882  
 QY 1172 gGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleAlaAsnVa 1192  
 DB 881 -----GGATATTAGTGTGAAAGCAT-----CTCGGTAT 853  
 QY 1192 IglNArgArgTrpArgAlaThrLeuGln-----MetArgArgGluArgLysAs 1208  
 DB 852 CCAGCATGACGACGCGTACAAATTTCAGTCTCACTTCGAAACACCCGCGCAGAGAAAGTC 793  
 QY 1208 nTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLysArgGl 1228  
 DB 792 CTACAGGAGGCTTCAGTGGCTGCTCAGTTTACAGACGTTACCGATCCCAACAGC 733  
 QY 1228 uMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValValGlnLy 1248  
 DB 732 AATGAAACAGGAA-----ATGCTGCTCTCTGAACGC 703  
 QY 1248 sArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGl 1268  
 DB 702 CAAGAAAGAGCTGTAATC----- 684  
 QY 1268 uValThrIleLysLeuGlnArgPheHisAlaGlnLysSerMetArgPheMetArgAl 1288  
 DB 683 -----ATTTACAGC 673  
 QY 1288 aLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsnHisLe 1308  
 DB 672 TGCCTTCCTGGA-----ATGAAATGCAGAGTCAATT 640  
 QY 1308 uLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGl 1328  
 DB 639 AAACGAAGG-----CACCAGCGCGCGGTATCATCCA 607

QY 1328 nArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLy 1348  
 DB 606 GAGAGTTTATAGAGCCGC-----CGTGAACGACAGCAGTATCTCATTGTAG 559  
 QY 1348 sGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnVa 1368  
 DB 558 ATCTCTGTGTGTGATCCATCCAGCGAGATATCCAGCCACTGTGTGCTGCAAAAGAGAAGT 499  
 QY 1368 lValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArgTyrArgGlyAs 1388  
 DB 498 AAAATCTTACCAACGATGCGCAGCAGCAGCTCTCTGTACAGCGCCCTTCAGGGGT-- 441  
 QY 1388 nLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLe 1408  
 DB 440 -----CAGCGGCTGAGGAAGAGTGGCGGCTGGCATCGAGCTGCACCTGTG--AT 391  
 QY 1408 uGlnLysTrpTrpArgSerIleArgAspMetArgLysLysAlaGlyTyrArgArgI 1428  
 DB 390 CCAGTCTGTGATTCAGAAAGCCACCCAG-----CAGATCAAGTTCAGGCCAT 343  
 QY 1428 eArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGl 1448  
 DB 342 GCGTCTGTCGCTGGTTCATTCAGAGATACACGCTCTCTGCATTCCTCAGAGCGGGA 283  
 QY 1448 nArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAl 1468  
 DB 282 GAGAGAGACCTTCCTGGAACGAGACGTTCTGCCATCGCTTCAGGAGCTTCAGAGG 223  
 QY 1468 aThrLeuLeuMetArgGlnGlnArgGluPheGluMetLysArgAlaAlaValVa 1488  
 DB 222 T-----CAGCGGCTGCGCGCAGTGTCAACAACATGCACAGAGTGCCTCT 175  
 QY 1488 lIleGln-----ArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAs 1505  
 DB 174 CATCAAGCAACATTCAGAGATTTAAGCTGCGCTCTGCATTTAGAGAGCAGCGCCAG-- 117  
 QY 1505 pTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSe 1525  
 DB 116 -----GCTGCTGTGTCTCCAGCAGAGGTTTCAGAGCCCAAGACA 76  
 QY 1525 rMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGl 1545  
 DB 75 GAAGAATGTGAGGTTAAACATTAACAGCTCAAGAAATGCTGCTGCTGTGTCAGGC 16  
 QY 1545 nLysPheArg 1548  
 DB 15 TGCATATCGC 6  
 RESULT 12  
 LOCUS AL720781/c  
 DEFINITION AL720781 Danio rerio embryonic inner ear subtracted cDNA Danio  
 rerio cDNA clone BNOA0472D10 5', mRNA sequence.  
 ACCESSION AL720781  
 VERSION AL720781.1 GI:20185385  
 KEYWORDS EST  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 526)  
 AUTHORS Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M.,  
 Hardelin,J.P., Weissenbach,J. and Petit,C.  
 TITLE A subtracted cDNA library from the zebrafish (Danio rerio)  
 embryonic inner ear  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).  
 Location/Qualifiers  
 1..526  
 source

/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="EN0AA047ZD10"  
/tissue\_type="inner ear"  
/dev\_stage="embryonic"  
/clone\_lib="Danio rerio embryonic inner ear subtracted  
cDNA"  
/note="subtracted cDNA library"

# ORIGIN

Alignment Scores: 2.54e-20 Length: 526  
Pred. No.: 359.50 Matches: 70  
Score: 359.50  
Percent Similarity: 67.50% Conservative: 38  
Best Local Similarity: 43.75% Mismatches: 49  
Query Match: 3.78% Indels: 3  
DB: 9 Gaps: 2

US-09-914-698-1 (1-1861) x AL720781 (1-526)

QY 645 PheAsnProLeuTrpLeuArgLeuGlyLeuGluValPheGlyGluLysIleGlnMet 664  
DB 504 TACAATCCACTTGGCTGGAAATGGACTGGACCAATTTATGGAGAGCTAATATCACTG 445  
QY 665 GlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArgAsn 684  
DB 444 GAAAGCAACAATGATGATGAGGACTGGCCATGTTATTCTTGGCGCTGCTGCTGTGGAAC 385  
QY 685 Lys-----CysGluGluGlnArgTyrSerLysAla---TyrThrLeuThrGluGluTyr 701  
DB 384 CCAGATTTAGCCGCTGAGTTGACACACTCTAAAGTGGCGCATTTGTATAGAGATGGCCAT 325  
QY 702 AlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuProPheLeuAsp 721  
DB 324 GAGGAGCACTGCTCGCTTACAGCAAGAACTAATCTGTTGGTCTCTTCTTGAC 265  
QY 722 GlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerPro 741  
DB 264 AAAGCTAAAGAACCCAGACTGATGATGAACTGATCTTGTGTTGTCATGGATGCTGAA 205  
QY 742 HisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGly 761  
DB 204 TTTAAGTCAAGCAAGATCTGCTGCTGCTTCTCAGAGACTTCCCTCAGTGTGAAGAT 145  
QY 762 AspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeu 781  
DB 144 ATTCTGCTCGGCACTCACTGCTGGGCTCGAGATATCCAGACTTCCCAACTCCCTT 85  
QY 782 AspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeu 801  
DB 84 GATGAGTTCAACTTTCGGGTGAAAAAATCGCGTTGACTTGAATGTGGCATCCGTTTA 25

# RESULT 13

AL720782/c  
LOCUS 526 bp mRNA linear EST 18-APR-2002  
DEFINITION Danio rerio embryonic inner ear subtracted cDNA Danio  
rerio cDNA clone EN0AA047ZD10 5', mRNA sequence.

AL720782 GI:20185386

VERSION EST.

KEYWORDS Danio rerio (zebrafish)

SOURCE Danio rerio

ORGANISM Danio rerio

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 526)

Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M.,

Hardelin,J.P., Weissenbach,J. and Petit,C.

A subtracted cDNA library from the zebrafish (Danio rerio)

embryonic inner ear

Unpublished (2002)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

# FEATURES

source

1..526  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="EN0AA047ZD10"  
/tissue\_type="inner ear"  
/dev\_stage="embryonic"  
/clone\_lib="Danio rerio embryonic inner ear subtracted  
cDNA"  
/note="subtracted cDNA library"

# ORIGIN

Alignment Scores: 2.54e-20 Length: 526  
Pred. No.: 359.50 Matches: 70  
Score: 359.50  
Percent Similarity: 67.50% Conservative: 38  
Best Local Similarity: 43.75% Mismatches: 49  
Query Match: 3.78% Indels: 3  
DB: 9 Gaps: 2

US-09-914-698-1 (1-1861) x AL720782 (1-526)

QY 645 PheAsnProLeuTrpLeuArgLeuGlyLeuGluValPheGlyGluLysIleGlnMet 664  
DB 504 TACAATCCACTTGGCTGGAAATGGACTGGACCAATTTATGGAGAGCTAATATCACTG 445  
QY 665 GlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArgAsn 684  
DB 444 GAAAGCAACAATGATGATGAGGACTGGCCATGTTATTCTTGGCGCTGCTGCTGTGGAAC 385  
QY 685 Lys-----CysGluGluGlnArgTyrSerLysAla---TyrThrLeuThrGluGluTyr 701  
DB 384 CCAGATTTAGCCGCTGAGTTGACACACTCTAAAGTGGCGCATTTGTATAGAGATGGCCAT 325  
QY 702 AlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuProPheLeuAsp 721  
DB 324 GAGGAGCACTGCTCGCTTACAGCAAGAACTAATCTGTTGGTCTCTTCTTGAC 265  
QY 722 GlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerPro 741  
DB 264 AAAGCTAAAGAACCCAGACTGATGATGAACTGATCTTGTGTTGTCATGGATGCTGAA 205  
QY 742 HisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGly 761  
DB 204 TTTAAGTCAAGCAAGATCTGCTGCTGCTTCTCAGAGACTTCCCTCAGTGTGAAGAT 145  
QY 762 AspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeu 781  
DB 144 ATTCTGCTCGGCACTCACTGCTGGGCTCGAGATATCCAGACTTCCCAACTCCCTT 85  
QY 782 AspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeu 801  
DB 84 GATGAGTTCAACTTTCGGGTGAAAAAATCGCGTTGACTTGAATGTGGCATCCGTTTA 25

# RESULT 14

CC129703/c

LOCUS 619 bp DNA linear GSS 16-APR-2003

DEFINITION ND.L19P17.SP6 Notre Dame Liverpool Aedes aegypti genomic clone

ND.L19P17, genomic survey sequence.

CC129703

CC129703.1 GI:29998758

GSS.

Aedes aegypti (yellow fever mosquito)

Aedes aegypti

Bukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;

Stegomyia.

1 (bases 1 to 619)

Loftus,B., Shetty,J., Knudson,D. and Severson,D.





```
Db      792 AAACAGGAAATGCTGCTCTGTAACGCCAAGAAAAAGA-----GCTGTAATCATC 745
QY      1300 LeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGln 1319
Db      744 TTACAGCTCTTTCCGTGGA-----ATGAAATGCAGAGTCAATTAAACGA 697
QY      1320 LeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLys 1339
Db      696 AGGCACCAGCCGCCGATGTCATCCAGAGAGTTTATAGACCCGC-----CGT 649
QY      1340 GlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArg 1359
Db      648 GAACGCAAGCAGATTCATCTTTGAGATCTCTGTGTGACCATCCAGCGGAGATATCGA 589
QY      1360 AlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIle 1379
Db      588 GCCACTGTGGCTCAAAAGAGAGTAAGTAAACTCTACCAACGGATGCCAGAGCAGCAGTC 529
QY      1380 LysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGln 1399
Db      528 CTGCTACAGGCAGCCTTCAGGGT-----CAGCGCGGTGAGGAAAGAGTGGCGCGC 478
QY      1400 LysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg 1419
Db      477 TGGCATCGAGCTGCAACTGTG---ATCCAGTCTGCATTCAGAACCCACAGCAG----- 427
QY      1420 LeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrp 1439
Db      426 -----CAGATCAAGTTCAGGCCATGCGTCTGCTCGCTGTGGTCAATTCAGAGATACTAC 373
QY      1440 ArgAlaThrValGlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysVal 1459
Db      372 CGCTCTCTGCAATCTCCAGAGCGCGGAGAGAGACCTTCCTGGAAACGAGACGTTCTGCC 313
QY      1460 ArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPhe 1479
Db      312 ATCGTCTCTTCAGGCAGCTTTCAGAGT-----CAGCGCGTGGCGCGCAGGTGC 265
QY      1480 GluMetLysArgArgAlaAlaValIleGln-----ArgArgPheArgAlaArg 1496
Db      264 AACAACTGCACAGAGTGCCTGCTGCTATCCAGCCCACTTCAGAGATTTAAGCTCAG 205
QY      1497 CysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuVal 1516
Db      204 TCTGCATTTAGGAGCAGCGCCAG-----GCTGCTGTGTCTCTC 166
QY      1517 GlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeu 1536
Db      165 CAGCAGAGGTTCAGAGCCCAAGACAGAGAATGTGGAGGTTAAACATTTACCACAGCTC 106
QY      1537 ArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGluGln 1556
Db      105 AGAAATGCTGCTGTCGTGTTGAGGCTGCATATCGTGAATGAAA----- 61
QY      1557 ArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGly 1576
Db      60 -----TGCAGAAATATCATCAAA----- 43
QY      1577 PheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArg 1596
Db      42 -----CGG 40
QY      1597 GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArg 1609
Db      39 AGGCATCGGCTGCAAGTGTATTATTCAGCGGCTTTTAGA 1
```

Search completed: September 15, 2004, 09:09:40  
Job time : 9641 secs

Alignment Scores:



[illegible]

3552	DB	TTATGGTGCATGTGCCCTAAGCCTGGACACAAAGAGAAACAGACTCCACCACACAGCAGCAACG	3611
1225	QY	-----AlaLysArgGluMetLysLysGlnArgAlaG	1235
3612	DB	AGTCTGAAATATCTTTAGCTCTGAAATTCGAGAAATGGAGACATTCCTCAGGACAG	3671
1235	QY	luPheLeuGlnLeuLysLysValThr-----LeuValValGlnLysA	1249
3672	DB	AGGAACCAAGTGAGAGAAGGTACTCTGGACATGTCAATTGCTTCTTAAGTCTCCAGAAGC	3731
1249	QY	rgArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluV	1269
3732	DB	GGGTCAACAGCTCGAGAGAGAGAGTGTATGTCAGATGAGCTGGACCGCAGGAGAGG	3791
1269	QY	alThrIleLysLeuGlnArgPheHisAlaGlnLysSerMetArgPheMetArgAlaL	1289
3792	DB	AGCAGGTGCTC-----CGCAGCAGGCCAAGGAGAGAAAGACCAC	3833
1289	QY	ysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsnHisLeu-	1308
3834	DB	AAATTAGAGGTGCAGAACTGGAATATGAGTCACCTCAAG-----CGTCAAGAACCTAG	3884
1309	QY	-----LeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrL	1327
3885	DB	AATCAGAAAAACAAAAACTGAAGATGAGCTAAATGATGTTCGCAAGGCC-----C	3935
1327	QY	euGlnArgArgTyrArgAlaArgLeuAsn-----MetI	1338
3936	DB	TCAGTGAGAAAGTGCCTCCAGAGGTGACCGCCCGCAGGTGCACCTGCCTACTGTGCTCTCA	3995
1338	QY	leLysGlnLeuLysSerTyrAlaGln-----LeuLysGlnAlaAlaIleThrI	1354
3996	DB	TGGAGCAGCTGACCTCTGTGAGCGAGGAGCTTGATGTCCGCAAGGAGGATCTCTCATCT	4055
1354	QY	leGlnThrArgTyrArgAlaLysLys-----Alam	1364
4056	DB	TAAGGCTTCAACTGGTCAGCCGAGAAAGAGCCATCCAAACCAAGGATGACAGAATACAA	4115
1364	QY	etGlnLysGlnValValLeu-----TyrGlnLysGlnArg-----	1375
4116	DB	TCACAGATTCCACATATCTTTTGAAGATGTACAAAAAATGAAGATAAAGGTGAATATAG	4175
1376	QY	--GluAlaIleIleLysValGlnArgArgTyrArgGlyAsnLeu-----GluM	1391
4176	DB	CACAAGATCATATTGGTTTGAGAAACAAATAGATCATCTGCTCTGGATTACCATGAGT	4235
1391	QY	etArgLysGlnIleGlu-----ValTyrGlnLysGlnArgGlnAla-----	1404
4236	DB	TGAATGAGGATGGAGAGCTGTGGCTGGTTTATGAAGGGTTTAAACCAAGCCATATGGCTCC	4295
1405	QY	--ValIleArgLeuGlnLysTrpArgSerIleArgAsp-----	1417
4296	DB	TGGAAATCCAGCTGCATCTCAGAGAGGAGCCATGAGATGAGGCGCGGCCCTCCGCTG	4355
1418	QY	-----MetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerI	1435
4356	DB	GGGAGATCCAGAGCCTGAAGAGGAGAGAACCAACCGACAGCAGCAGCTGCTGGCCCCAGAAC	4415
1435	QY	leGlnArgLysTrpArgAlaThrValGlnAlaArgGlnArgGluIlePheLeuSerT	1455
4416	DB	TGCAGCTGCCCCCGAGAGCCCGATGTTGAGCGCCAGCCTGCACGACGAGATCACCCTGGGTGA	4475
1455	QY	hrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnG	1475
4476	DB	CCAACGAAACTTGGATTGATGAAACA-----CITGAAARACAGG	4517
1475	QY	lnArgArgGluPheGluMetLysArgArgAlaAlaValValIleGlnArgPheArgA	1495
4518	DB	ATAAGCGGTCCGTAAACTGAAAAAACAACCTGAAAGTATTTGGCAAAAAATTTGGCGAAC	4577
1495	QY	laArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerVal----	1513
4578	DB	TAGAAGTGGCCACGATGGAGAAATATATCCCGAGACAGATCATTTGATGAACCATCCGAC	4637

Qy	1514	--lleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaAArgGlnGluP	1533
Db	4638	CAGTCAACATTCCGAGGAAAGAAAGATTCCAAAGGATGCTGGAATACAAGAAGAGG	4697
Qy	1533	heValGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuM	1553
Db	4698	ATGAGCAAAACATTGTTTAAAGAACCTGATTCTGGAACCTGAAGCCACGCTGGTGTACAGTCA	4757
Qy	1553	etIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaA	1573
Db	4758	ATTG-----ATTCAGGATTACCGGCAT	4781
Qy	1573	rgAlaArgGlyPheMetAlaArg-----LysArgPheGlnA	1585
Db	4782	ATAFCTGTTCATGTGTTCGACATGCTGACTACCTGAATGATGATCAAGAAATAAGGT	4841
Qy	1585	laLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleG	1605
Db	4842	CGTGTCTAAACATCAACAATTAAACAGCATC-----AAAAAGTATTGA	4883
Qy	1605	lnArg-----TyrTrpArgGlyTyrLeuIleArg-----	1614
Db	4884	AGAAAAGAGGTGATGATTTCAAACCTCTCCTTCTGGCTCTTAACACATGCGGATTT	4943
Qy	1615	-----ArgArgG	1617
Db	4944	TGCACCTGCTTGAACACAGTACAGTGGAGAGGGCTTTATGAAGACACACACATCTCGCC	5003
Qy	1617	lnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnAlaL	1637
Db	5004	AGAATGACACTGCCTCACCATTCT-----GACCTGGCTGAGTATCGGCAG-----	5050
Qy	1637	ysAlaValAsnSerValArgCysLysVal---GlnGluAlaValArgPheLeuArgGly-	1655
Db	5051	--GTGCTGAGTGACTTGGCCATTGCAGCTACCAGCAGCTCGTGGCGGTCTGTAGAGAACA	5108
Qy	1656	-----ArgPheIleAlaSerAspAlaLeu-----	1663
Db	5109	TCCTTCAGCCAATGATTGTCTCAGCGCATGTGTGAACATGAACAGATTACGGGCGTGTCTG	5168
Qy	1664	-----AlaValLeuSerGlnLeuAspArgLeuSerArgT	1675
Db	5169	GGGTGAAGCCACAGGGTTGAGAAGCGAACCTCCAGTATCGCCGACAGGGGCACCTACA	5228
Qy	1675	hrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleM	1695
Db	5229	CACGTGACTCCATCCTCGGCAGCTCAACTCCTTCCACTCGGTCAATGTGTGAGCATGGCA	5288
Qy	1695	etAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgIleI	1715
Db	5289	TGGACCTTGAACTGATCAAGCAGGTGTGCAAGCAGATGTTCTACATCATAGGGGCCATCA	5348
Qy	1715	leLeuAsn-----LeuAlaA	1720
Db	5349	CCCTGAACAACCTTCTCTCTGCGGAGGACATGTGCTCTCTGGAGTAAAGGCATGCAGATCA	5408
Qy	1720	rgTyrAsnSerThrThrVal-----AsnThrPheGlnGlnGlyG	1733
Db	5409	GGTACAATGTCACTCACTGAAAGAAATGGTGTGCGTGACAAGATCTGATGAATAGTGGG	5468
Qy	1733	ly-----LeuValThrIleAlaGlnMetLeu-----LeuArgTrpC	1745
Db	5469	CTAAGAAACCCCTGGAACTCTCAATCAGGCTGTCTCAACTTTCGAGTGAAGAAAGAAA	5528
Qy	1745	ysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisC	1765
Db	5529	CAGATGATGATGCAGAA-----GCCATTCTGTCTAT-GTGCAATGCTTTAACTACT	5578
Qy	1765	ysPro-----LysLysA	1769
Db	5579	GCCAGATGTGTGAAGTGTTGAAATTTGTATCTCCAGTTAATGAGTTTGAGAAAGAGTC	5638

```

Qy 1769 rgfysllelleHisAspTyrMetThrAsn 1778
      :::::||||| ||| |||:::
Db 5639 TCTGTGCTGCTTACCTCGTACTATACAGAT 5667

RESULT 2
US-08-056-200-93
; Sequence 93, Application US/080556200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fredrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1645..2511
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2512..8070
; US-08-056-200-93

Alignment Scores:
Pred. No.: 3,85e-17 Length: 9551
Score: 294.00 Matches: 323
Percent Similarity: 36.08% Conservative: 291
Best Local Similarity: 18.98% Mismatches: 637
Query Match: 3.09% Indels: 453
DB: 1 Gaps: 66

US-09-914-698-1 (1-1861) X US-08-056-200-93 (1-9551)

```



Db 5652 GGAGGCAATATCCGAGAGGAGGAGCTGCAGCAGCAGGAGGAGCAGCTGCCTCAGAGA 5711  
 Qy 990 ualaalalePheLeuGlnArgGlieTrpArgArgArgLeuPheAlaLysLysLeuLeuAl 1010  
 Db 5712 GGAACCGGAGAGAGAGCCCGCAGGACCGGAGGAGCAATGTCGGGAGGAGGAGGAGCT 5771  
 Qy 1010 aAlaAleuThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAlaSerTyrIl 1030  
 Db 5772 GCAGCAGGAGAGAGCAGCTGCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGCT 5831  
 Qy 1030 eGlnMetGlnTrpArg---ThrTyrGlnLeuGlyArgGlieGlnArgHisGluPheLeuAr 1049  
 Db 5832 GGAGAGCAATATCCGAGGAGGAGGAGGAGCTTCAGCGCCAGAGGAGGAGGAGGAGGAG 5891  
 Qy 1049 gGlnArgAspLeuMetPheValGlnArgMetArgSerLysTyrTrpSerMetLeuG1 1069  
 Db 5892 GGATGAGAT-----CAGCGCAGTATCGAATGCGAGTGGGAACC 5933  
 Qy 1069 uGlnArg-----LysG1 1073  
 Db 5934 AGAAAAAGAAATCGAGTTCGTGATAACAAGGTTTACTGCAAGGCGCAGAGAGAATGAACA 5993  
 Qy 1073 uPheGlnGlnLysArgAlaAlaIleAsnIleGlnGln----- 1086  
 Db 5994 GTTCCGCGAGTGGAGATTCCCGAGTCCCGAGTCCCGACAGACAGATCCCGAGCAAGATCGCAGCA 6053  
 Qy 1087 -----ArgTrpArg----- 1089  
 Db 6054 CCTGCTGGTGAACAGCAGAGAGAGATCGTGACAGAGAGGAGGCGCTGCGCAGCAGGC 6113  
 Qy 1090 -----AlaLysLe 1092  
 Db 6114 CAACAGGCAATTCAGAGGAGAAACACAGCTGGAGCGAGAGAGCAAAAGGAGCCAAAG 6173  
 Qy 1092 uSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValG1 1112  
 Db 6174 GCGCGACAGAGTCCCAAGAGGAGAAACAGATTCTGAGAGAGGAGGAGGAGGAGGAGAG 6233  
 Qy 1112 nAlaTyrArgLysAlaThrIleGlnMetArgGlieAspArgAsnHisTyrTyrSerLeuAr 1132  
 Db 6234 ACCCGCTCAGAGACAGACAGAGAAATTCGCGAGGAGGAGAAACAGCTGCTCCAGGAAAGGGA 6293  
 Qy 1132 gLysAsnValIleCysLeuGlnArg----- 1141  
 Db 6294 GGAACAGCCGCTGCTCGCCAGAGCGTGACAGAGAAATTCGCGAGAGGAGAACTGCTCCA 6353  
 Qy 1142 -----LeuArgAlaIleMetLysMetArgGlnArgGlu-- 1153  
 Db 6354 TCAGGACAGGAGAGAAATTCCTCGAGGAGGAGAACAGCGCTGCGCAGGAGGAGGAGAG 6413  
 Qy 1154 -AsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgG1 1173  
 Db 6414 AAAATTCCTTAAGAGGAGAACAG-----CAGCTGCGCTCGAGGAGCGCGA 6458  
 Qy 1173 nGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValG1 1193  
 Db 6459 GCAACTGCGTCCAGGAGCGGACAGAGAAATTCGCGAGGAGGAGAACAGCAGCTGAGCGGCCA 6518  
 Qy 1193 nArgArgTyrArg-----AlaThrLeuGlnMetArgGluAr 1206  
 Db 6519 AGAGCGTGACAGAAATTCCTGTAAGAGGAGAACAGCGGTGCGCCGCGGAGGAGGAGAG 6578  
 Qy 1206 gLysAsnTyrLeuHisLeuGlnThrThrThrLysArgGlieGlnIleLysPheArgAlaLys 1226  
 Db 6579 AAAATTCCTGAGGAGGAGAACAGCAGCTGCGCAGGAGGAGCTCAGAGAAATTCGCGGAGAG 6638  
 Qy 1226 sArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLysValThrLeuValVa 1246  
 Db 6639 GGAACAGCTGCTCCAGGAGGAGGAGAA---CAGCAGCTGCACCGC----- 6681  
 Qy 1246 lGlnLysArgArgArgAlaLeuLeuGlnMet----- 1256

Db 6682 -CAAGAGCGTGACAGAGAAATTCCTGGAGGAGGAGAAACAACAGCTGCGCCGCCAAGAGCGTGA 6740  
 Qy 1257 -ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeu----- 1273  
 Db 6741 CAGAAATTCGCGAACACAGAGAAATTCGCGAGTCCAGGAAACACAGAGAGAAATTCCTCGAGGA 6800  
 Qy 1274 -GlnArgArgPheHisAlaGlnLysSerMetArg-----PheMe 1286  
 Db 6801 GGAACAGCAGCTGACCGCCAGCAACCGCAGAGAAATTCCTCCAGGAGGAGAACAGCAGCT 6860  
 Qy 1286 tArgAlaLysTyrArgGlyThrGln-----AlaAlaVa 1297  
 Db 6861 GCGCGCGCAGGAGCGCGGCAACAGCGCGTCCAGCAGCGTCCAGAGAAATTCCTCGCGAGGA 6920  
 Qy 1297 lSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPh 1317  
 Db 6921 GGAACAGCTGCGCGCAGGAGGAGGAGAACAGCAGCTGAGCGCGCAGAGCGTGCAGAGAA 6980  
 Qy 1317 eLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgTyrArgAlaArgLeuAsnMe 1337  
 Db 6981 ATTCCGTTAGAGGACAGAGAAATTCGCGCCAGAGAACAGAGAGAAATTCATGGAGGA 7040  
 Qy 1337 tIleLysGlnLys-----SerTyrAlaGlnLysGln----- 1349  
 Db 7041 CGAACAGCAGCTGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7100  
 Qy 1350 -----AlaAlaIleThrIleGlnThrArgTyrAr 1359  
 Db 7101 CCGCGAAGACGAAACAGCTGCTCCAGGAAAGGAGAAACAGCAGCTGCAGCCGCAAGAGCG 7160  
 Qy 1359 gAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIl 1379  
 Db 7161 TGACAGAAATTCCTCGAGGAGGAGAACCGCAGCTGCGCGCCAGGAGGAGGAGGAGGAG 7215  
 Qy 1379 eLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrG1 1399  
 Db 7216 -----CAGCTGCGTCCAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7268  
 Qy 1399 nLysGlnArgGlnAlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetAr 1419  
 Db 7269 GGAAGGAGGAGAACAGCAGCTGCGCGCCAGAGCGGTGACAGAAATTCGCGGAGAGGAGGA 7328  
 Qy 1419 gLeuCysLysAlaGlyTyrArgArgIleArg-----LeuSe 1431  
 Db 7329 ACAGCAGCTCCGCGCTCAGAACAGAGAGAGAAATTCCTCCAGGAGGAGAACAGCAGCTGCG 7388  
 Qy 1431 rSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIl 1451  
 Db 7389 CCGCCAGGAACTCGAGAGAAATTCGCTGAGGAGGAGAACAGCTGCGCCAGAGGAGGAG-- 7446  
 Qy 1451 ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLe 1471  
 Db 7447 ----CAGAGCAGCTGCGCGCCAGAACAGACGCTACAGAGAAATTCCTAGAGGAGAGCAGCT 7502  
 Qy 1471 uMetArgGlnArgArgGluPheGluMetLysArgArgAlaAlaValValIleGlnAr 1491  
 Db 7503 C---CGTCCGAGGAGGAGAACAGCAGCTGCGCGCCAGGAGCGC-----GACAG 7550  
 Qy 1491 gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeu---IleG1 1510  
 Db 7551 AAAATTCGCGAGGAGGAGAACAGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7610  
 Qy 1510 nSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaAr 1530  
 Db 7611 AGAGTCT-----GACAGAAATTCGCGAGGAGGAGAACAGCTACGCCAGGAGAG 7658  
 Qy 1530 gGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLy 1550  
 Db 7659 GGAAGAA---CAGCAGCTGCGCCCCAACAGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 7715  
 Qy 1550 sArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPh 1570  
 Db 7716 GCAGCTCCAACTTGAGGAGGAGGAG-----CAGAGGCTGCGG----- 7752







Db 5994 GTTCCGGCAGTTGGAAGATTCCAGGTCGCGGACAGACAATCCAGCAGAAATCTGCAGCA 6053  
QY 1087 -----ArgTrpArg----- 1089  
Db 6054 CTTCTGGTGAACAGCAAGAGAGAGATCGTGAGCAAGAGAGAGGCGCTGCGCAGCAGGC 6113  
QY 1090 -----AlaLysLe 1092  
Db 6114 CAACAGGCATTTCCAGAGAGAACACAGCTGAGCGGAGAGCAAGCAAAAGGAGCCAAAG 6173  
QY 1092 uSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGI 1112  
Db 6174 GCGGCACAGGAAGTCCCAAGAGGAAAAGACGTGCTGAGAGAGGAAAAGAGAGAGAG 6233  
QY 1112 nAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuAr 1132  
Db 6234 ACGCGCTCAGAGACAGACAGAAAATTCGCGAGGAGGAACAGCTGCTCCAGGAAAAGGGA 6293  
QY 1132 gLysAsnValIleCysLeuGlnArg----- 1141  
Db 6294 GGAACAGCGCTGCTCCGCAAGAGCGTGACAGAAAATTCGCGAAGAGGAACGTGCTCCA 6353  
QY 1142 -----LeuArgAlaIleMetLysMetArgGlnArgGlu-- 1153  
Db 6354 TCAGGAACAGGAGAGAAAATTCCTCGAGAGGAACAGCGGCTGCGCGAGGAACGGGAG 6413  
QY 1154 -AsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGI 1173  
Db 6414 AAAATTCCTTAAGAGGAAACAG-----CAGCTGCGCTCCGAGGAGCGCGA 6458  
QY 1173 nGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGI 1193  
Db 6459 GCAACTGGCTCAGACCGCGACAGAAAATTCGCGAGGAGGAACAGCAGCTGAGCGCGCA 6518  
QY 1193 nArgArgTrpArg-----AlaThrLeuGlnMetArgGluAr 1206  
Db 6519 ACAGCGTGACAGAAAATTCCTGTAAGAGGAACAGCAGCGTGCGCGCCAGGAACGAGAGAG 6578  
QY 1206 gLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLys 1226  
Db 6579 AAAATTCCTGAGGAGGAACAGCAGCTGCGCCAGGAGCGTCCACAGAAAATTCGCGGAAGA 6638  
QY 1226 sArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVa 1246  
Db 6639 GGAACAGCTGCTCAGGAAGAGGAAGAA--CAGCAGCTGCACCGC----- 6681  
QY 1246 lGlnLysArgArgAlaLeuLeuGlnMet----- 1256  
Db 6682 -CAAGAGCGTGACAGAAAATTCCTGGAGGAGGAACAACAGCTGCGCGCCCAAGAGCGTGA 6740  
QY 1257 -ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeu----- 1273  
Db 6741 CAGAAAATTCGCGAACAAGAACACGCGAGTCCAGGAACCCAGAGAAAATTCCTCGAGGA 6800  
QY 1274 -GlnArgArgPheHisAlaGlnLysSerMetArg-----PheMe 1286  
Db 6801 GGAACAGCAGCTGCACCCGCAAGCAACGCGCAGAGAAAATTCCTCCAGGAGGAACAGCAGCT 6860  
QY 1286 tArgAlaLysTyrArgGlyThrGln-----AlaAlaVa 1297  
Db 6861 GCGCGCCAGGAGCGCGGCAACAGCGCGTCCAGGAGGAACAGAAAATTCGCGGAGGA 6920  
QY 1297 lSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPh 1317  
Db 6921 GGAACAGCTGCGCCAGGAGAGGAGGAACAGCAGCTGAGCGCGCCCAAGAGCGTGACAGAAA 6980  
QY 1317 eLeuGlnLeuArgAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMe 1337  
Db 6981 ATTCCGTTAGAGNACAGAAAGTGCAGCGCCGAGGAACAAGAGAAAATTCATCGGAGGA 7040  
QY 1337 tIleLysGlnLeuLys-----SerTyrAlaGlnLeuLysGln----- 1349  
Db 7041 CGAACAGCAGCTGCGCGCCAGGAGGGGCCAACACAGCTGCGCCAGGAGGAGCAGAAAATTT 7100

QY 1350 -----AlaAlaIleThrIleGlnThrArgTyrAr 1359  
Db 7101 CCGCGAAGACGAACAGCTGCTCCAGGAAGAAAGAACAGCAGCTGCACCGCCAGAGCGG 7160  
QY 1359 gAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleI 1379  
Db 7161 TGACAGAAAATTCCTCGAGGAGGAACCGCAGCTGCGCGCCAGGAGCGGAACAA----- 7215  
QY 1379 eLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGI 1399  
Db 7216 -----CAGCTGCTGTCAGCAGCGCGACAGAAAATTCGTCGAGAGGAACAGCTGCTCCA 7268  
QY 1399 nLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetAr 1419  
Db 7269 GGAAGGGAGGAACAGCAGCTGCGCGCCCAAGAGCGTGACAGAAAATTCGCGGAGAGGA 7328  
QY 1419 gLeuCysLysAlaGlyTyrArgArgIleArg-----LeuSe 1431  
Db 7329 ACAGCAGCTCGCGCTCAGGAACGAGAGAAAATTCCTCCAGGAGGAACAGCAGCTGCG 7388  
QY 1431 rSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGlnArgGluI 1451  
Db 7389 CCGCCAGGAACCTGAGAGAAAATTCGTCGAGGAGGAACAGCTGCGCCCAAGAAAACGGAG-- 7446  
QY 1451 ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLe 1471  
Db 7447 ---CAAGAGCAGCTGCGCGCCCAAGAACGCTACAGAAAATTCCTAGAGGAGAGCAGCT 7502  
QY 1471 uMetArgGlnGlnArgArgGluPheGluMetLysArgArgAlaAlaValIleGlnAr 1491  
Db 7503 C---CGTCCGGAAGGAGGAACAGCAGCTGCGCGCCAGGAGCGC-----GACAG 7550  
QY 1491 gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeu---IleGI 1510  
Db 7551 AAAATTCGCGAGGAGGAACAGCTCCGCGCAGGGAAGGAGGAACAGCAGCTGCGCAGCCA 7610  
QY 1510 nSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaAr 1530  
Db 7611 AGAGTCT-----GACAGAAAATTCGCGAGGAGGAACAGCTTACCGCCAGGAGAG 7658  
QY 1530 gGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnLysPheArgGlyLy 1550  
Db 7659 GGAAGAA---CAGCAGCTGCGCCCCAACACGCTGACGGAAGTATCGCTGGGAAGAAGA 7715  
QY 1550 sArgLeuMetIleGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPh 1570  
Db 7716 GCAGTCCCACTTGAGGAACAAGAG-----CAGAGGCTGCGG----- 7752  
QY 1570 eGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGI 1590  
Db 7753 -CAGAGCGAGACCGGAGTACCGGCGGAGGAGCAGGTTGCCACG----- 7797  
QY 1590 uMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGI 1610  
Db 7798 -----CAGGAGAGAGTCTGTCGAGGAACAAGAACTATGGCAAGA 7838  
QY 1610 yTyrLeuIleArgArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAl 1630  
Db 7839 AGAGGAGCAGAAAACGTCGCCAGGAACGGGAAGAAAATTCGCGGAAGAAACATCCCGCGCG 7898  
QY 1630 aGlnLeuArgGlnGlnAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaVa 1650  
Db 7899 CCAGAGAGAGGAACAGAGGCAACCGCAAGTCGCGGAGGATACATCCCAAGAAAGGAA 7958  
QY 1650 l-----ArgPheLeuArg-----GlyArgPheIleAla-SerAspAlaLeuA 1664  
Db 7959 GGGCCATGGGCGGCTTCTGGAGCCCGGCACATCATGTTGCCAGTGTCGCCAGTGGGCTC 8018  
QY 1664 laValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1684  
Db 8019 CAGCCCTCTCTATGAGTACATCCCAAGAGCAGAGATCTCAATACCGCCCTTAAGTATGTT 8078







```

; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Dmanac, Radole T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784C1F2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)...(8082)
US-09-620-312D-130

Alignment Scores:
Pred. No.: 3,19e-15 Length: 8503
Score: 273.50 Matches: 382
Percent Similarity: 36.20% Conservative: 383
Best Local Similarity: 18.08% Mismatches: 761
Query Match: 2.87% Indels: 587
DB: 94 Gaps: 94

US-09-914-698-1 (1-1861) x US-09-620-312D-130 (1-8503)
QY 107 ProProGlnAlaProLeuValGluLysAsnValTyrLysThrProGlnGluProVal 126
Db 2347 CCTCTGAGTGAAGAGGCTGAGGAAAGAGATACAGACAAATCTGAAGAGCTCCATATA 2406
QY 127 TyrIleSerProGlnProArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsn 146
Db 2407 ATAACATCAGAAAAGATATAATGTTTCTGAA----- 2439
QY 147 LeuLeuAspValIleAspAsnLeuArgPheThrProLeuThrCluThrArgGlyLysGly 166
Db 2440 -----GTAGTTTCATAGGAGAGATGAGTTCAAGGTTTACTTGAAGAAATTGGGAAACA 2493
QY 167 GlnAlaThrIlePheProAspAsnLeuAlaAlaTyrProThrProThrLeuLysGlyAsn 186
Db 2494 AAA-----GATGACCTAGCA-----ACTACACAGTCGAAT 2523
QY 187 ValLysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuGluAsp 206
Db 2524 TATAAAGC-----ACTGATCAAGAAATTCCTCAAAAT 2553
QY 207 GlnProAlaThrAsnLysThrPheAspValLysHis----- 218
Db 2554 TTCAAAACCTTCATATGAGCTTTGACCAAGAAAGTATAAGATGGTCTTGAGGAGATGAG 2613
QY 219 -----SerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAsp 235
Db 2614 AGAATGAATCAGGAAATAGTAACTCTCTAAGAACCCCAAAATTTGATCGAGTTG 2673
QY 236 GlyGlnProHisThrProLeuAsnLysThrThrThrIleValHisAlaThrHisThrArg 255
Db 2674 GGTGCTTTGAAGCCGAGCTTCTTACAGAACCCCAAGAACTTCAGGAGAAAACACGTGAG 2733
QY 256 AlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThrProThrLysSer 275

```

```

Db 2734 GTTCAAGAAAAGACTAAATGAGATCGAAACAGCTGAAGGAACAATTAGAAAATAGAGATTCT 2793
QY 276 AlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuAlaGlySer 295
Db 2794 CCGCTGCAAACTGTAGAAAGGAGAAAACACTGATTACTGAGAAACCTGCACAACTTTA 2853
QY 296 GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySer 315
Db 2854 GAAGAAGTAAAAAACTTAACT----- 2874
QY 316 MetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlu 335
Db 2875 -----CAAGAAAAGAGATGATCTAAACAACACTCCAAGAAAGCTTCCAATTTAGAGG 2925
QY 336 GlnGlnIleGlnIleLysAlaLysAsp-----LeuAsnSerSerSerSerSerGluAlaSer 354
Db 2926 GACCAACTCAAAAGTGATATTCACGATCTCTTAACTATATATAGTACTACAGAACAA 2985
QY 355 LeuAlaGlyGlnGlnGluPheLeuPheAsnHis-----Ser 366
Db 2986 TTACGAATGCTCTTGAGTCTCTGAAACACATCAAGAAACAATTTATATACACTAAATCG 3045
QY 367 GluIleLeuAlaGlnSerSerArgPheAsnLeuHis----- 378
Db 3046 AAAATTTCTGAGGAAGTTTCCAGG---AATTTGCATATCGAGGAAAATACAGAGAAACT 3102
QY 379 -----GluValGlyArgLysSerValLysGlySerProValLysAsnProHisLysArg 396
Db 3103 AAAGATGAATTTTCAGCAAAAGATGTT---GCCATAGATAAAAAACAGGATTTGAGACT 3159
QY 397 ArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGlu 416
Db 3160 AAAAATACCCAAACACTAACTGCAGATGTTAAGGATATAGATATAATTGAGCACAAGG 3219
QY 417 ThrVal-----AlaIleSer 421
Db 3220 AAGATATTTCTTTAATACAGGAGAAAATGAATCCAACTCCAAATGTTAGAGATGTTATA 3279
QY 422 ProProLysLysGln-----ArgValGluAsp 430
Db 3280 GCAGAAAAGGAAACAAATGGAAGACTGACCTTAAGGAAAATATTGAAATGACCTTGAAC 3339
QY 431 ThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHis 450
Db 3340 CAGGAAGAATTAGACTTCTTGGGATGAACCTTAAAGAACCAACAGAGATAGTTGCACAA 3399
QY 451 AlaTyrProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLys 470
Db 3400 GAAAAGAACCATGCCATATAAAGAAAGAGAGAGACTTTCTAGGACCTGTGACAGACTGGCA 3459
QY 471 LysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyr 490
Db 3460 GAA-----GTTGAAGAAAACCTAAAGAAAGAGCCAGCCCACTCCCAAGAAA----- 3507
QY 491 AspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThr 510
Db 3508 -----CAGCAACAACCTTCTAAT----- 3525
QY 511 IleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGln 530
Db 3526 -----GTACAAGAGAGATG 3540
QY 531 AlaAspPheLysLysThrLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAsp 550
Db 3541 AGTGAATGCGAGAAAAGATTAATGAATA-----GAGAAATTAAGAATGAA 3588
QY 551 LeuAsnAsnLys-----IleAspValGlyLys 559
Db 3589 TTAAGAACCAAGAAATTGACATTTGGAACATATGGAACACAGAGAGGCTTGAGTTGGCTCAG 3648
QY 560 LeuPheAsnGluValArgAsnLysGluLeuValAlaProThrLysGluGluGlnSer 579

```



Db 3649 AAACCTTAATGAA-----AATTATGAGGAAGTGAAATCTATAACCAAGAAAGAAAGTT 3702  
QY 580 MetAsnTyrLeuThrLys-----  
Db 3703 CTAAGGAATATACAGAAGTCATTTGAAACAGAGAGAGACCACCTTAGAGGATATATAAGA 3762  
QY 586 -----TyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu 598  
Db 3763 GAAATTGAAGCTACAGGCTCAAAACCAAGAAAGAACTAAAAATTGCTCATATTACCTA 3822  
QY 599 -----PhePheSerGluGlnMetArgLeuProCysSerLysValAlaVal 613  
Db 3823 AAAGAACACCAAGAACTATTGATGAACCTAAGAGAGCGTATCTGAGAGACAGCTCAA 3882  
QY 614 TyrValAsnLysGlnAlaLeuArgLysArgSerAspArgAsnLeuHisLeuAspValVal 633  
Db 3883 ATAATAAATACTCAGGACTTAGAA---AAATCCCATACCAATTTACAGAAGAGATCCCA 3939  
QY 634 MetGlnArgThrLysLeuGluLeuLeuLeuCysPheAsnProLeuTyrLeuArgLeuGly 653  
Db 3940 GTGCTTCATGAGGAACAAGAGTTACTG-----CCTAATGTGAAAAAAGTCAGT 3987  
QY 654 -----LeuGluValValPheGlyGluLysLysLysLysLysLysLysLysLys 665  
Db 3988 GAGACTCAGGAACAATGAATGAATGAGTTATTA-----ACAGAACAGTCC 4035  
QY 666 SerAsnArgAspLysValGlyLeuSerThrPheLysLeuAsnArgLeuPheArgAsnLys 685  
Db 4036 ACAACCAAGACTCAACACACACTGGCAAGATAGAAATGGAAGGCTCAGTTCAATGAA 4095  
QY 686 -----CysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTyrAlaGlu 703  
Db 4096 AAATTCACAGAAAGTCGGAAGAGATATAATCTCTACCAAGGAAAGAGACAACCTTAAA 4155  
QY 704 ThrLeuLys-----LysHisSerLeuGlnLysLysLysLysLysLysLysLysLys 718  
Db 4156 ACGATAAAGAGCCCTTGAGTTAAACAT-----  
QY 719 PheLeuAspGlnAlaLys-----GlnLysArgLys 728  
Db 4186 -----GACCAGCTGAAAGAACATATTAGAGAACTTTGCTAAATCCAGGAGTCTCAA 4239  
QY 729 ValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLysAspLys 748  
Db 4240 AGCAAAACAAGACAGTCCCTTAAATATGAAAGAAAGAAAGACAATGAAACTACCAAAATCGTG 4299  
QY 749 -----LeuLeuArgPheSerSerGluLeu 756  
Db 4300 AGTGAGATGGAGCAATTCAAACCAAGATTTCAGCTACTAAGGATAGAAATAGAAATG 4359  
QY 757 LeuAla-----AsnLysGlyAspLys 763  
Db 4360 CTGGGATTTGTCAAAAGACTTCAAGAAAGTCATGATGAATGAATCTGTAGCTAAGGAG 4419  
QY 764 ThrArgGluLeuArgArgLeuGlyTyrValLeuGlnPheArgGlnThrPheLeuAspGlu 783  
Db 4420 AAAGATGACCTACAGAGGCTGCAAGAAAGTTCTTCAA-----TCTGAA 4461  
QY 784 PheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArg 803  
Db 4462 AGTGACCACTCAAAGAAACATATAAGAAATTTAGCTAAACACCTGGAAACTGAAGAG 4521  
QY 804 ValValGluVal-----LysLeuLeuArgAspAspLeuThrArgGlnLeuArg 819  
Db 4522 GAACTTTAAAGTTTGCTCATTTGTTGCTGGAAGAAACAAGAGGAACTATTATGATTTAAGA 4581  
QY 820 ValProAlaLysSerArgLeuGlnArgLysPheAsnValLysLeuAlaLeuGlyAlaLeu 839  
Db 4582 GTGAATCTTTCAGAGAGAACTGAATATCAACCATTCAAAGCAGTTAGAACCAATC 4641  
QY 840 GlyGluAlaAsnPheGlnLeuGlyGlyAspLysAlaAlaGlnAspValAspGlyHis 859  
Db 4642 -----AATGATAAATTACAGAACAGATC-----CAAGAGATTATGAGAAAGAG 4686

QY 860 ArgGlu-----LysThrLeuSerLeuLeuTyrGlnLeuLeuLysPheArgSer 876  
Db 4687 GAACAACTTAATAATAAAACAAATTAGTGAGGTTTCAGGAAAAAGTGAATGAATGAACAA 4746  
QY 877 ProLysPheHisAlaAlaAla-----ThrValLeu 886  
Db 4747 TTCAGGAGCATCGCAAGCCCAAGGATTCAGCACTCAAAAGTATAGAAGTAAGATGCTC 4806  
QY 887 GlnLysTyrTrpTrpArgHisTrpLeuHisValValLysGlnArgArgLysHisLys 906  
Db 4807 GAGTTGACCAACAGACTTCAAGAAAGTCAAGAGAAATACAAATATGATTAAAGGAAAAA 4866  
QY 907 GlnLeuMetArgHisArgAlaAlaThrValLysGlnAlaValPheArgGlyHisGln 926  
Db 4867 GAGGAATGAAAGAGTACAGGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAAA 4926  
QY 927 MetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaLysLysLys 946  
Db 4927 ACTAAAGAAATGTAGCTAAATGAAAGATCTCAAGAAAGAAATATCAGTTCTTT 4983  
QY 947 LysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerLysLys 966  
Db 4984 AAGATGACAGTGTCAATGAGACTCAGGAGAAATGTGTGAAATAGAACAC 5034  
QY 967 ThrLysGlnArgTyrTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVal 986  
Db 5035 -----TTGAGGAGCAATTTGAGACCCAGAGTTAAACCTGGAACACATAGAAACG 5088  
QY 987 GlnLeuArgGluAlaAlaLysPheLeuGlnArgLysLysLysLysLysLysLysLys 1006  
Db 5089 AATATAAGGTTGACTCAGATACATGATAAACTTCAAGAAATGAGATCTGTATCAACAAA 5148  
QY 1007 Lys-----LeuLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGln 1021  
Db 5149 GAAAGAGATGACCTTAGAGTGTGAGAGACTCTCAAAGTAGAGAGAGCCAGCTCAAG 5208  
QY 1022 -----LysGlnGlnAlaAlaLysTyrLys 1030  
Db 5209 GAAACCTTAGAGAACTATACTAGACCTAGAAACCAAGAGGAGCTAAAAATTTGT 5268  
QY 1031 GlnMetGlnTrpArgThrTyrGln-----HisGluPheLeuArg 1049  
Db 5269 CACATGCACTGGAAGGAGCACCACCAAACTATTGTATAAACTAAGAGGATTTTTCAGAG 5328  
QY 1039 -----LeuGlyArgLysGlnArg----- 1049  
Db 5329 AAACAAATGAAATATCAATATGCAAAAGGACTTAGAACACCTCAAAATGATGCTTAAAA 5388  
QY 1050 GlnArgAspLysLysMetPheValGlnArgArgMetArg---SerLysTrpSerMetLeu 1068  
Db 5389 GCACAGATCTG-----AAATATCAAGAGGAACTAAGAAATGCTCACATGCAATCGAAA 5442  
QY 1069 GluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaLysLysLysLysLysLys 1088  
Db 5443 GAGCAGCAGGAACTATTGACAAACTCAGA-----GGAATTTGTTTCTGAGAAG 5490  
QY 1089 ArgAlaLysLeuSer---MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSer 1107  
Db 5491 ACAGATAAACTATCAATATGCAAAA-----GATTTAGAAATTTCA 5532  
QY 1108 ValLeuLysValGlnAlaTyrArgLysAlaThrLysGlnMetArgLysPheArgAsnHis 1127  
Db 5533 AATGCTAAATTAACA-----GAAAGATTCAGAACTTAAGGCAATGAAATCAACAA 5583  
QY 1128 TyrTyrSerLeuArgLysAsnValLysCysLeuGlnGlnArgLeuArgAlaLysMetLys 1147  
Db 5584 CTTATTACCTTAAAAAAGATGCTCAATGAGACACAGAAAAAGTGTCTGAAATGGAGCAA 5643  
QY 1148 MetArgGluGln---ArgGluAsnTyrLeuArgLeuArgAsnAlaSerLys 1163  
Db 5644 CTAAGAAACAAATAAAGAACCAAGCTTAACCTCTGAGTAAATTAGAAATAGAGATTTTA 5703

Qy 1164 ---LeuValGlnLysArgGlyTyrArgMetArgGlnGlnMet-----IleGlnAspArg 1179  
Db 5704 AATTGGCTCAAGAACTTCATGAAGAACTTCGTAATGAAAGAAAGA 5763  
Qy 1180 AsnAlaTyrLeuArgThrArgLysCysIle----- 1189  
Db 5764 GATAATCTAAGAGAGTAGAGAGACACTCAAACTGGAGAGAGACCAACTCAAGGAAAGC 5823  
Qy 1190 -----IleAsnValGlnArgArgGlyTyrArgAlaThrLeuGln 1201  
Db 5824 CTGCAAGAAACCAAGCTVAGAGATCTGGAATACACAGCAACTAAACACTGCTGATG 5883  
Qy 1202 MetArgArgGluArgLys-----AsnTyrIleHisLeuGlnThrThrLysArg 1218  
Db 5884 CTATCAAGAAACACAAAGAACTGTTGATAAACTTAGAGAAATAATTTCAGAAAGACA 5943  
Qy 1219 IleGlnIleLysPheArgAlaLys----- 1226  
Db 5944 ATTCAATTTTCAGACATTCAAAAGGATTAGATAAATCAAAAGATGAATTACAGAAAAAG 6003  
Qy 1227 ---ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuVal 1245  
Db 6004 ATCCAAGAACTTCAGAAAGAACTTCAACTGCTTAGAGTCAAAAGAGATGTCATATG 6063  
Qy 1246 ValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGln----- 1261  
Db 6064 AGTCATAAAAAAATAATGAATGGAACAGTTGAAAGAGCAATTTGAGCCAAAACACTATCTA 6123  
Qy 1262 -----GluTyrLeuHis-----LeuArgGlu 1268  
Db 6124 TGCAAGCTGAGATGATAAATCCAGTTGACTAAGAAACCTTCATGAAGCCCTTGAGAA 6183  
Qy 1269 ValThrIleLysLeuGlnArgArgPheHisAlaGlnLys-----SerMetArgPhe 1285  
Db 6184 ATAAGAATTTAGCTAAAGAAAGAGATGAGCTAAGGAGATAAAAGAACTCTCAAAATG 6243  
Qy 1286 MetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeuGlnMetHisTrpArg 1305  
Db 6244 GAAAGGGAACCAATTCATAGCAACCTTAAGGGAATG-----ATAGCTAGAGACCGACAG 6297  
Qy 1306 AsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaLys 1325  
Db 6298 AACCAACCAAGTAAACCTGAAAAAGG-----TTACTAAGTGTGACACACAG 6345  
Qy 1326 ThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGln-----LeuLysSerTyr 1344  
Db 6346 CACCTTATGGAAGCCCTGAGAGAAAGTCTCTAGATAAAAGAGCTTTTGAAGAGATAC 6405  
Qy 1345 AlaGln-----LeuLysGlnAlaAlaIleThrIleGlnThrArg 1357  
Db 6406 TCAGAGATGGATGATCATATGAGTGTGATGATGATGCTCTCTGACTTGGAGAGGNA 6465  
Qy 1358 TyrArgAlaLysLysAlaMetGln-----LysGlnValValLeuTyr----- 1371  
Db 6466 ATTTGAATCCACAGATCATGAAGAACTGAAGTATGTTGTAAGCTATGTTACAAAAATA 6525  
Qy 1372 ---GlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGlyAsnLeuGlu 1390  
Db 6526 AAAAGAGAACCAACATGATCCATCAATAAATTTGAAATGGATTT----- 6570  
Qy 1391 MetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLys 1410  
Db 6571 -----ATTGATGAAGTGGAAAGCAAGAAATGCTAATTAATAATACAGCAC 6618  
Qy 1411 TrpTrp-----ArgSerIleArgAspMetArgLysCysLysAla 1423  
Db 6619 CTTCAACAGATGTGATGATCATCCAGAGATTAAGGAATCTCAAAATGAACACAG--- 6675  
Qy 1424 GlyTyrArgArgIleArgLeuSerSerLeu----- 1433  
Db 6676 ---AATATGATCTACATATTAGGAAATTTCTCAAGATTTTCTCAGAAAGATGATGCTCCCT 6732  
Qy 1434 SerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGlnArgGluIlePheLeu 1453

Db 6733 AGCATAAGAGCTGAATTTCAACAAGTACTAAGTAATAGAAAGAAATGACACAGTTTTTG 6792  
Qy 1454 Ser-----ThrIleArgLysVal----- 1459  
Db 6793 GAAGAGTGGTTAAATACTCGTTTTGATATAGAAAGCTTAAATAATGGCATCCAGAAAGAA 6852  
Qy 1460 -----ArgLeuMetGlnAla-----PheIleArgAlaThrLeuLeuMetArgGln 1474  
Db 6853 AATGATAGATTTCCTCAAGTCAATAACTCTTTTAACACAGAAATAATGGCATAATGAAT 6912  
Qy 1475 GlnArgArgGluPheGluMetLysArgAlaAlaValIleGlnArgPheArg 1494  
Db 6913 GAATCAACAGAGTTTGAGAAAGA-----AGTGTACCATATCCAAAGAGTGGGAA 6963  
Qy 1495 AlaArgCysAlaMetLeuLysAlaArg-----GlnAspTyrGlnLeuIle 1509  
Db 6964 CAGGACCTGGAATCACTGAAAGAGAAAAATGAAAAACCTATTATAAACTCCAAACATTTG 7023  
Qy 1510 GlnSerSerValIle-LeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAl 1529  
Db 7024 AAGACTTCTTGCGATCTGGTCCAGGTTAATCTCTACCAACAAGACA----- 7072  
Qy 1529 aArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu----- 1543  
Db 7073 -----ATAAGAAATCTCTCATGTTACATCAAGAGCTACACAGTTAAACC 7113  
Qy 1544 -GlnGlnLysPheArgGlyLysArgLeuMetIleGlnArgAsnCysPheGlnLeuIle 1563  
Db 7114 ACAGAGAAATTCGAGAGCTGGAATAAT-----CACTGCATGAAGCTAAA 7158  
Qy 1563 uArg-----CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLys 1581  
Db 7159 GAAAGTGTATGCATA-----AGGAAGCAAGATTTATAAGATGACAGAA 7203  
Qy 1581 sArgPheGlnAlaLeuMetThr-----ProGluMetMetAspLeuIle 1595  
Db 7204 GAACCTTGAGTGACTTAATGACATAATGCAAACTTCAAGCCAAAGTTTCATGAATCAAT 7263  
Qy 1595 eArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIle-Arg- 1614  
Db 7264 AAATGCCCTTGAAAAACAAAGAGACAATCAAGTACTTCAGGACAAAGTTGCTTTAGGA 7323  
Qy 1615 ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeu----- 1632  
Db 7324 GCTAAGCCATATAAGAGAAATTCGAAGTCTCAAAATGAAGCTTGTGAATAGACCTA 7383  
Qy 1633 -----ArgGlnGluAlaLysAlaValAsnSerValArgCysLysVal 1646  
Db 7384 GAGAAATGAAAAATGCCAAAGAAATTTGAAAAAGAAATCAGTGTACAAAAAGCCACTGTA 7443  
Qy 1647 Gln-----GluAlaValArgPheLeuArgGlyArgPhe-----IleAla 1659  
Db 7444 GAATATCAAAAGGAGATTTATAGGCTATTGAGAGAAATCTCAGAGAAAGTCAACAGGCC 7503  
Qy 1660 SerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeu 1679  
Db 7504 CAAATACCTCAGTGTATCAGACATACATCTGATCTCAGCCTTCAATAAACC----- 7557  
Qy 1680 LeuMetTrpCysSerGlu-----PheMetSerThrPheCysTyrGlyIleMet 1695  
Db 7558 ---TTAACTTGTGGAGGTGGCAGGCGCATTTGACAAAAACACAAAGCTCTTTATTTGAAA 7614  
Qy 1696 AlaGlnAlaIleArgSerGluValAsp-----LysGlnLeu 1707  
Db 7615 AGTGAAACATATAAGCTAGAAAAAGAAATTTCTAAGTTAAACAGCAAAAATGAACAGCTA 7674  
Qy 1708 IleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsn 1727  
Db 7675 ATAAACAAAGAAATGAATTTGTAGCAAT---AATCAGCATCTTCCATGAGTCAAA 7731  
Qy 1728 ThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAspLys 1747

```

Db      7732 ACTTGAAGAA-----AGAACCCCTAAAGAGAGAGGCTCACAAA 7770
QY      1748 AspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysProLys 1767
Db      7771 CAAGTAACITGTGAGATTCT-----CCAAG 7797
QY      1768 LysArgLysIleLeuHisAspTyrMetThrAsnProGluAlaIleTyrMetValArgGlu 1787
Db      7798 TCTCCTAAAGTGAAGTGA-----ACAGCTTCT 7824
QY      1788 ThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaArgLysProPro 1807
Db      7825 AAAAGAAACAATACACCTCTCAATGCAAGACGGAATTTACAAAGATCCCTGTGCCA 7884
QY      1808 MetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrProCysSerLeuProSer 1827
Db      7885 AAGGAATCACCA-----AAATCTGTGTTTGTATAGCCGATCAAGATCTTTACCAATCA 7938
QY      1828 LeuGluProAspPheGlyIleLeuArgTyrSerProTyrThrPhe-IleSerSerValty 1847
Db      7939 CCTCATCCA-----GTTGCTAT-----TTTGATAACTCAAGTTTA 7974
QY      1847 rAlaPheAspThrIleLeuCysLysLeuGln 1857
Db      7975 GGCTTT--GTCAGAGGTGCAAAATGCAG 8002

RESULT 6
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Alignment Scores:
Pred. No.: 5.93e-12 Length: 580073
Score: 273.50 Matches: 304
Percent Similarity: 36.58% Conservative: 271
Best Local Similarity: 19.34% Mismatches: 598
Query Match: 2.87% Indels: 400
DB: 4 Gaps: 71

US-09-914-698-1 (1-1861) x US-08-545-528D-1 (1-580073)
QY      296 GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyr----- 310
Db      259797 GAGATCAAAAAGAGATTATTAAGTATTGAGTATGAAATACCTATATAGAGATGATAGT 259856
QY      311 -----AlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerIle 327
Db      259857 GCAGATAATCAACTGCAAAATGTTTATGAAACATGCATCAAAATCGATCCAGTTTAAA 259916
QY      328 GluGlnAsnArgTyrTyrGln-----GluGlnGlnIleGlnIleLys 341
Db      259917 CACCAATACCAAACTTATAGAGATGAGTTAAGTCAACTTGAAGCGCAAGATCCAGCTCACC 259976
QY      342 AlaLysAspLeuAsnSerSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPhe 361

```

```

Db      259977 AAACAAGAGTTAGTTGATAAAGAAATCAGCACTAAGAGTAAGAGATAGATGCTGTGATTT 260036
QY      362 LeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGly 381
Db      260037 TACATTAAATCCCGT-----TTAGCTGAA-----CTTGATGATGTAGCT 260075
QY      382 ArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGluLeu 401
Db      260076 AAA-----CAAACCTT 260084
QY      402 SerPheSerAsp-----AlaProSerAsnGluSerLeuTyrArgAsnGlnThrValAla 419
Db      260085 AGCTTTCAAGATGGTATTACCAAGCAAAATGCTCAACATCTTGAGGATAAGTTAGTTGCT 260144
QY      420 IleSerProLys-----LysGluArgValGluAspThrThrLeuProArg 435
Db      260145 TTAACACAAAGAAAAGACACCGTTTAAATACCCAAAGAGAGCGCTTTTAACTTAAAGACAA 260204
QY      436 SerAla---AlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHis 454
Db      260205 TCTGCTTTAATGATATCATATAAATCCAGCAGAAAATGAAGTGTGCT---AAGCAC 260261
QY      455 Ala-----GlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLys----- 470
Db      260262 TTAGAAACACCAAGCAAAATGAGTTTGAACAAAAACAGTCTGATAGCTCTTTAAAGCTTGAA 260321
QY      471 -----LysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLys 488
Db      260322 ACTGAATATAAGGCTTTTACAACCAAGATTAATGAATTTTAAATAATGAAGTGCACATAAG 260381
QY      489 LeuTyrAspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThr 508
Db      260382 -----AGCGAGGAACCTTTTAAACCAA-----GAACGGGAACCTATTTGAAAAACGC 260426
QY      509 ThrThrIleAspProPheLeu-----AlaSerThrMetTyrLeuAspGlu----- 523
Db      260427 AGGGAATTTGACACGCTTTTAAACCAAGCATCTTTGGAATATGAACACCAACGTGAGTCA 260486
QY      524 ---GlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSer 542
Db      260487 AGTCAACTACTCAAAGATAAGCAGATGAATGAATAAACAACACATCTCCAAACCTTAGAGTAT 260546
QY      543 IleProAlaAspLeuAspAlaAspLeuAsn-----AsnLysIleAspVal 557
Db      260547 GCTAAAAGAGGAGCTAGACAAAGAAAGAACCTCTTAGATCAACAAAAAAGTTGATAGT 260606
QY      558 GlyLysLeuPheAsnGluValArgAsnLysGluLeuValValAlaProThrLysGluGlu 577
Db      260607 GAAGCAATCTTT-----CAACTCAAGAAAAAGTTGCTCAAGAACCTAAGGAGCTT 260657
QY      578 GlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGlu 597
Db      260658 GAAGAACTA---TACCTTGTATAAAACAAAAACAAGATCAGAAGGAAAAATGAAGT--- 260711
QY      598 LeuPhePheSerGluGlnMetArg-----LeuProCys 608
Db      260712 TTGTTTTTGAAGAGCAGTTTAAACACCAACAGCAGATTTGAAATGAATGGAGCT 260771
QY      609 SerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSer----- 624
Db      260772 AAACAACAGGAGTTGTTTGAAGCTAAACATGATCATGGAA---CGTTCCTTTATCAAACTT 260828
QY      625 ---AspArgAsnLeuHisAspValValMetGlnArgThrIleLeuGluLeuLeuLeu 643
Db      260829 GAAGATTAAGAAAAAGATCTTAAACCTAAGCACAACAG----- 260867
QY      644 CysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlyLysIleGln 663
Db      260868 -----ATTGCCAATCAGTGTTCCTCCAA 260888
QY      664 MetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArg 693

```





Db 1426 ATAACAACAAAACACATAGCTTTCTATATAATTTTATACGAGAAATGATGATCTGTC 1485  
Qy 140 -----SerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThr 157  
Db 1486 TGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTAAGTGAATGAATGAAT 1545  
Qy 158 ProLeuThrGluThrArgGlyGlyGlnAlaThrIlePheProAspAsnLeuAla 177  
Db 1546 CCAGCAACAAAGCTACTAAATCAG-----GAGAATATAGAAGT 1584  
Qy 178 TrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArg 197  
Db 1585 ---GAGTTGAACCTACCTCGTCTGACTATGATATCTGGTATTAGACTATGAACAAC 1641  
Qy 198 ArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPheAsp 215  
Db 1642 CGAAGCAAGAAAGAAAGAAATGGAATTAATAAGAAAGAAAGATGATTGGATGAATTT 1701  
Qy 215 ----- 215  
Db 1702 GAGGCTCTAGAAAGAAACAACTAAAAAGATCAAGAGATGCAACTAATTCATGAATTTGCG 1761  
Qy 216 -----ValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp 230  
Db 1762 AACTTAAGATTTAGTTAAGCATCGAGACTATATATCAAGATCTTTGAGATGAATC 1821  
Qy 231 CysSerArgIleAsp-----GlyGlnProHisThrProLeuAsnLysThrThrIle 248  
Db 1822 AGTTCAAAAGTAGAGCTCTTAGAGAAAGAAAGAACACAGATTAAGAGACTACAGGAATAC 1881  
Qy 249 ValHisAlaThrHis-----ThrArgAlaLeuAlaCysIle 260  
Db 1882 ATAGACTCTCAAAAGCTAGAAAAATATAAAATGGAAGCTGTGCTATCATTCATGGAAGCA 1941  
Qy 261 HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280  
Db 1942 GAAGACCCAAACAAATGAACAGACTCTGTTGATGCTGAACTGTAGCCCTTGATGCC 2001  
Qy 281 LysArgAspIleLysLeuValGlySerProLeuArgLysThrSerGluSerMetLysAsp 300  
Db 2002 AAGAGAGAAATCAGCCCTTCTTAGAAGTCAAAATCTGGAGTTGAAGAGAAATGAAGAA 2061  
Qy 301 LeuSerLeuLeuSerProGlnThrLysThrAlaIleGln----- 313  
Db 2062 CTTCGAACATACATACAGCAAAATGGAATATGATATTCAGTTATATCAAGCCCAATGGAG 2121  
Qy 313 ----- 313  
Db 2122 GCMAAAAGAAATGCAAGTTGATCTGGAGAAAGAAATTAACAATCTGCTTTTAATGAGATA 2181  
Qy 313 ----- 313  
Db 2182 ACAAAACTCACCTCCTTATAGATGGCAAGTTCCAAAAGATTTGCTCTGTAATTTGGAA 2241  
Qy 314 -----GlySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArg 331  
Db 2242 TTGGAAGGAAGATTAAGTCTCTCAGAAAGAACTAATAAAGAAAGTTGAAGAAATGAA 2301  
Qy 332 TyrTyrGlnGluGlnGlnIle-----GlnIleLysAla----- 342  
Db 2302 GCTTTGCGGGAAGAGTCAATTTGCTTTTTCAGAAATTTGAATCTTTTACCTCTGAGTAGAA 2361  
Qy 343 -----LysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGln 359  
Db 2362 AGGCTGAGGAAAGAGATACAGACAAATCTGAAGAGCTCCATATAATAACATACAGAAAAA 2421  
Qy 360 GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsn-----Leu 377  
Db 2422 GATAAATTTGTTT-----TCTGAAGTAGTTTCAAGAGAGAGTAGGTTCAAGGTTTACTTT 2475  
Qy 378 HisGluValGlyArgLysSerValLysGlySerProValLysAsnProHisLysArg 397  
Db 2476 GAAGAAATTTGGGAAACAAACAAAGATGACCTAGCACTACAGCTCGAATTTATAAAGCACT 2535

Qy 398 SerHisGluLeu---SerPheSerAspAlaProSerAsn---GluSerLeuTyrArgAsn 415  
Db 2536 GATCAAGATTCCAAATTTCAAACCCCTTCATATGAGCTTTTGACGCAAAAGTATAAG--- 2592  
Qy 416 GluThrValAlaIleSerProLysLysGlnArgValGluAspThrThrLeuProArg 435  
Db 2593 -----ATGGTCTCTGAGGAGAAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAA 2646  
Qy 436 SerAlaAlaProAlaAsnAlaSer-----AlaArgSerSerSerAlaHisAlaTrp 452  
Db 2647 GAAGCCCAAAATTTGATTCGAGTTTGGTCTTTGAGACCGAGCTTTCTTCAAGACC 2706  
Qy 453 ProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysPro 472  
Db 2707 CAAGAACTTCAGGAGAAACACAGTGAGGTTCAAGAAAGACTAAATGATGGAACAGCTG 2766  
Qy 473 AlaThrProArgLysValArgAspThrSerIleGln-----ProSerValLysLeuTyr 490  
Db 2767 AAGGACAAATTAGAAATAGAGATTCTCGCTGCAAACTGTAGAAAGGGAGAAACACTG 2826  
Qy 491 AspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThr 510  
Db 2827 ATTACTGAGAACTGCAGCAAACT----- 2850  
Qy 511 IleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGln 530  
Db 2851 -----TTAGAAGATGATTAATAACTTTTAACCTCAAGAA 2880  
Qy 531 AlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAsp---LeuAspAla 549  
Db 2881 AAGATGATCTAAACAACTCCAGAAAGCTTGCAAAITTGAGAGGCACTCAAAAGT 2940  
Qy 550 AspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeu 569  
Db 2941 GATATTCAAGTACTGTTTAACTG---AATATAGATCTCAAGAACTAATAATCGAAATTTCTGAG 2997  
Qy 570 ValValAlaProThrLysGluGluGlnSerMetAsnTyrLeu---ThrLysTyrArgLeu 588  
Db 2998 CTTGAGTCTCGAAACATCAAGAAACAAATTAATACACTAAATCGAAATTTCTGAG 3057  
Qy 589 GluThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys 608  
Db 3058 GAAGTTTCCAGGAATTTGCATATGAGGAAATACAGGAGAAACTAAAGATGAATTTCTCAG 3117  
Qy 609 SerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeu 628  
Db 3118 CAAAGAGATGGTGGCATAGATAAAAAACAGGATTTTGAAGCTTAAATAATACCCAAACACTA 3177  
Qy 629 HisLeuAspVal-----ValMetGlnArgThrIleLeuGluLeuLeu 642  
Db 3178 ACTGCAGATTTAGGATTAAGATAATGAGTAATGAGCAACAAAGAGATATTTCTTTAATA 3237  
Qy 643 LeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlyLysIle 662  
Db 3238 CAGGAGAAAAATGAA-----CTCAACAATGTTAGAGAGTGTATAGCAGAAAAAGGAA 3291  
Qy 663 GlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsn----- 679  
Db 3292 CAATTGAAGACTGACCTAAAG---GAAAAATATTGAATTCACCATTTGAAAAACAGGAGAA 3348  
Qy 680 ---ArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThr 698  
Db 3349 TTAAGACTCTTGGGATGAACCTTAAAGCA----- 3381  
Qy 699 GluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuPro 718  
Db 3382 CAAGAGATGTTGCACAGAAAGAACCATGCCATAAAGAAAGAGAGGAGGAGCTTTCTAGG 3441  
Qy 719 PheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLys 738  
Db 3442 ACCTGTGAC-----AGACTGGCAGAAAGTTGAAGAAAAACTAAAGGAAAAAG 3486







Db 5373 AAATGATGCCTTAAAGACACAGGATCTGAAATATACAGAGGAACTAAGATTGCTCAT 5432  
Qy 1389 uGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaVal----- 1405  
Db 5433 GCATCTGAAGAGCAGCAGGAACTATTGACAACTCAGAGGAATTGTTCTGGAAGAC 5492  
Qy 1406 -----1LeArgLeuGlnLysTr 1411  
Db 5493 AGATAAACTATCAATATGCAAAAGATTTAGAAAATTCAAATGCTAAATACAGAA-- 5550  
Qy 1411 pTrpArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArgAlaIleArgLeuSe 1431  
Db 5551 -----AAGATTCAAGAACTTAAGGCAATCAACATCACTTATTACGTTTAAAAAAGA 5603  
Qy 1431 rSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluI 1451  
Db 5604 TGCAATGAGACACAGAAAAGTGTGAAATGGAGCACTAAAGAAACAAATAAAGA 5663  
Qy 1451 ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeu 1471  
Db 5664 CCBAAGCTTAACCTGAGTAATAGAAATAGAGAAATTTAAATTTGGCTCAAGAACTCA 5723  
Qy 1471 uMetArgGlnGlnArgGluPheGluMetLysArgAlaValAlaValIleGlnAr 1491  
Db 5724 TGAATAACCTTGAAGAAATGAAATCTGTAATGAAAGAAAGAGATATCTA-----AG 5774  
Qy 1491 gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSe 1511  
Db 5775 AAGAGTAGAG-----GAGACACTCAAACTGGAGAGAGAC---CAACTCAAGGAAAG 5822  
Qy 1511 rSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgG 1531  
Db 5823 CCTCGAAGAACCAAGCTAGAGACTCGAATACACAGGAACTAAATACTGCTCGTAT 5882  
Qy 1531 nGluPheValGln-LeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGly---- 1549  
Db 5883 GCTATCAAAAGAACACAAAGAACTGTTGTAATACTTAGAGAAAAATTTTCAGAAAGAC 5942  
Qy 1550 -----LysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuL 1563  
Db 5943 AATTCAAATTCAGACATTCAAAGAGATTTAGATAAATCAAAAG----- 5986  
Qy 1563 euArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgP 1583  
Db 5987 -----ATGTAATACAGAAAGAAAGATCCAGAACTTC---AGAAAAAGAACT 6029  
Qy 1583 heGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLys-----ArgAlaA 1601  
Db 6030 TCAACTGCTTAGA-GTGAAGAGAGATGTCAATATGAGTCATAAAAAAATTAATGAAATGG 6088  
Qy 1601 laLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArg----- 1614  
Db 6089 AACAGTTGAAGAGCAATTTGAGCCAACTATCTATGCAAGTGTCAGATGGATTAATCTCC 6148  
Qy 1615 -----ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuA 1633  
Db 6149 AGTTGACTAAGAACTTCATGAAGCCCTTGAAGAAATAAGA---ATTGCTAGCTAAGAAA 6205  
Qy 1633 rGlnGlnAlaLysAlaValAsnSerValArgCysLysValGlnGlnAlaValArgPheL 1653  
Db 6206 GAGATGAGCTTAAG-----AGGATAAAGAAATCTCTCAAAATGG 6244  
Qy 1653 euArgGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArg---- 1671  
Db 6245 AAAGGACCAATTCATGCAACC-----TTAAGGCAATGATAGCTAGAGACCGACAGA 6298  
Qy 1672 -----LeuSerArgThrValProHisLeuLeuMetT 1682  
Db 6299 ACCACCAAGTAAACCTGAAAAAGGTTACTAAGTATGGACACAGACCTTATG----- 6354  
Qy 1682 rpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerg 1702  
Db 6354 ----- 6354

Qy 1702 luValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrA 1722  
Db 6355 -----GAAAGCCTGAGAGAAAGTGCTCTAGAATAAAGAGCTTTTGAAGAGTACT 6406  
Qy 1722 snSerThrThrValAsnThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuL 1742  
Db 6407 CAGAG-----ATGGATGATCATTATGATGCTTGAATAGATTGTCTC 6448  
Qy 1742 euArgTyrCysAspLysSerGluIlePheAsnThrLeuCysThrLeuIleTrpValP 1762  
Db 6449 TTGAC---TTGGAGAGGAAATTTGAATTTCCACAGAATCATGAAGAAACTGAATATGTGT 6505  
Qy 1762 heAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaI 1782  
Db 6506 TAACTATGTTTACAAAATAAAGAAAGCAACATGATGATCATATAAATTTGAAATGG 6565  
Qy 1782 leTyrMet-----ValArgGluThrLysLysLeuValAlaArgLysGlnLysMetLysG 1800  
Db 6566 ATTTTATTTGATGAAGTGGAAAAAGCAAAAGGAATTTGCTAATTTAAAAATACAGACCTTCAAC 6625  
Qy 1800 lnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsn- 1819  
Db 6626 AAGATTGTGATGTACCA-----TCCAGAGAAATTAAGGATCTCAAAATTTGAACC 6673  
Qy 1820 -----PheThrProCysSerLeuProS 1827  
Db 6674 AGAATATGATCTACATATTTAGGAAATTTCTCAAGATTTCTCAGAAAGTGAATCCCTA 6733  
Qy 1827 erLeuGluProAspPhe 1832  
Db 6734 GCATAAAGACTGAATTT 6750  
RESULT 8  
US-08-328-354-5  
; Sequence 5, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8789 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 544..7990
; US-08-328-254-5

Alignment Scores:
Pred. No.: 1,17e-14 Length: 8789
Score: 268.00 Matches: 367
Percent Similarity: 34.89% Conservative: 327
Best Local Similarity: 18.45% Mismatches: 747
Query Match: 2.82% Indels: 552
DB: 1 Gaps: 85

US-09-914-698-1 (1-1861) x US-08-328-254-5 (1-8789)
QY 105 HisProProGlnAlaProLeuValGluLysAsnValTyrIlys----- 119
Db 2776 CACCCTGTGCTTCTTGGCTCCATTGGACGAGAGTAATCTTACGAGCACTTGACATTTGTCA 2835
QY 120 -----ThrProGlnGluProValTyrIleSerPro 130
Db 2836 GACAAAGAGTTCAAATGCACCTTTGCCAATTGCAAGAGAAATCTTATCTTTACAAAGT 2895
QY 131 GlnProArgSerLeuLysGluAsnLeuSerProMetThrPro-----GlyAsnLeuLeu 148
Db 2896 GAACACAAATTTTACATGATCAGACTGTCAGATGAGCTCTAAATGTCAGAGCTGCGAG 2955
QY 149 AspValIleAspAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlnAla 168
Db 2956 ACCTAGTTGACTCANTTA-----AAGGCCGAAAT 2985
QY 169 ThrIlePheProAspAsnLeuAlaIaTrpThrProThrLeuLys----- 184
Db 2986 TTGGCTTTGTACGAATCTGAAACTTTCAAGGTGACTTGTGTAAGGAGATGCGAGCTG 3045
QY 185 -----GlyAsnValIysSerCysAlaAsnAspMetArgProArgArgIleThr 200
Db 3046 GGCCTTGGAGGAGGGCTGCTTCCATCCTCTGTCATCTCTTGTGCTGCTGACAGCTCTAGT 3105
QY 201 ProAspAspLeuIleAsp-----GlnProAlaThrAsnLysThrPheAspVal 216
Db 3106 CTTAGCAGTTTGGGAGACTCCTCTTTTACAGAGCTCTTTTGAACACAGACAGAGATATG 3165
QY 217 LysHisSerGluThrIle-----AsnIleSerLeuAspThrLeu 229
Db 3166 TCTCTTTGAGTAATTTAGAGGGGCTGTTTTCAGCAACAGTGCAGTGTGATGATGAAGTA 3225
QY 230 AspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal 249
Db 3226 TTTTTCAGCAGCTGCGAGGAGAGATCTGACC----- 3258
QY 250 HisAlaThrHisThrArgAlaIleAlaCysIleHisGluGluGluGlyProSer-ProPr 269
Db 3259 -----AGGAAGAAACCCCTTCGGCCCCA 3282
QY 269 oArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySe 289
Db 3283 GCGAAGGTGTTGAAGAGC-----TTGATCCCTCTGTGAGGTGTACCGGCAG 3330
QY 289 rProLeuArgLysTyrSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLy 309
Db 3331 TCCCTC-GAGAAGCTAGAGAGAAATGGAAGT----- 3363
QY 309 sTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArsSerIleGlu 329
Db 3364 -----CAGGGATTATGAATAAT-----AAGGAAATTCAGAGCTCGAGCA 3404
QY 329 n-----AsnArgTyrTyrGlnGluGlnG 337
Db 3405 GTTATTAAAGTCTGAAGGCAAGAGCTTGACTGCCCTTAGAAGCAGTATTTGTCAAAAA 3464

```

```

QY 337 nIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaG 357
Db 3465 TGAACAGTGGCAACAGACAGCTGACAGCTCTGGAGATGAGTCCCAAGTTGGCGGC 3524
QY 357 yGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLe 377
Db 3525 AGAAAAGAAA-----CAGACGGAACAACGTCTCACT 3554
QY 377 uHis-----GluValGlyArgLysSerValLysGlySerProValLysAsnProHisLysAr 396
Db 3555 TGAGCTGGGAAGTAGACAGCTCCAGCTCAAGATCTGGACTTAAAGTTCT-----CG 3605
QY 396 gArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnG 416
Db 3606 GTCTTTGCTTGGCATCGACACAGAGATGCTTATCAAGC-----CGAAATGA 3653
QY 416 uThrValAlaIleSerProLysLysGlnArgValGluAspThrThrLeuProArgSe 436
Db 3654 GAGCTGTGACATATCA-----AAAGAACATATCTTCAGAACTACCA-----GA 3695
QY 436 xAlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaG 456
Db 3696 AAGAACACCAAGCATGATGTCATCAGATTGTGTATAAGATGCTCAGACGAGACCTCAA 3755
QY 456 nSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysPro-----Al 473
Db 3756 TCTAGACATTGAGAAATAACTGAGACTGGTGCAITG-----AAACCCACAGGAGAGTG 3809
QY 473 aThrProArgLysValArgAspThrSerIleGlnProSerValLysLeuLysSerG 493
Db 3810 CTCTGGGGAACAGTCCCCAGATACCAATATGAGCCTCCAGGGGAA-----GATAAAAAC 3863
QY 493 uLeuTyrMetGlnThrCysIleAsnProAspPropheAlaAlaThrThrThrIleAspPr 513
Db 3864 CCAGGCTCTTCAGATGCAATTCGCAATTCATTCCTGCTCTTAAGCTTTGTGTACC 3923
QY 513 o-----PheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGlnAl 531
Db 3924 TATGGATTTCTCTGGG-----AATCAGGAAGATATCCATAATCTTCAACT 3968
QY 531 aAspPheLysLysTrpLeuAsn-----AlaLeuValSerIleProAlaAspLeuAs 548
Db 3969 GCGGGTAAAGACATCAATCAGAAATTTGAGATTAATCTCATGTATGATAGAGACCGTGA 4028
QY 548 pAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLysG 568
Db 4029 CAGA-----AAAGTTGAAAGTTTGCTAAATGAATG----- 4059
QY 568 uLeuValValAlaProThrLysGluGlnSerMetAsnTyrLeuThrLysTyrArgLe 588
Db 4060 -----AAAGAATTAGACTCAAACTCCATTTACAGGAGGTACAACCT 4100
QY 588 uGlnThrLeuArgLysAlaAlaValGluLeu----- 598
Db 4101 AATGACCAAAATGGAAGCATGATGATGATAATGAAAAAATAGTTGGGAACTTAAGAAAGA 4160
QY 599 -----PhePheSerGluGlnMetArg--LeuProCysSerLysValAlaValTyrVa 615
Db 4161 AAACCTCAGATTTAAGTGAATAANTGGAATATTTCTGTGATCCACGAGAGTACTCCA 4220
QY 615 lAsnLysGlnAlaLeuArg---IleArgSerAspArgAsnLeuHisLeuAspValValMe 634
Db 4221 CAGAGTAGAAACTTCTGAAGGCTCAATCTCTGATTTAGAAATGTCATGACAGATAAATCATC 4280
QY 634 tGlnArgThrIleLeuGluLeuLeuLysCysPheAsnProLeuTrpLeuArgLeuGlyLe 654
Db 4281 ACGTGAAGATATTGGAGATAATGTGCCCAAGGTGAATGACAGCTGAAGAGAGAGATTCT 4340
QY 654 uGluVal-----ValPheGlyGly 661
Db 4341 TGAATGGGAATAGCTGAGTAGGATCAGATCGGAGAAAGCTAGCATTCAGCATGAAGC 4400
QY 661 sIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLe 691

```

Db 4401 CCTCTACCTGGAGCTGATCTTAGAGGTAGTTCAAACACAGAGAGCTATGTTTATAGAAAAGA 4460  
Qy 681 uPheArgAsnLys 690  
Db 4461 CAATGAAATTAAGCAGAGAGGTTATTGTTCTGCTTGAAGAGAACTCTCAGTGGTGCACAAAG 4520  
Qy 691 699  
Db 4521 TGAGAGAAACACAGCTTCTGTGAGAGATTAAGATATGTTCAAAAAAACCACCGGCACTGGA 4580  
Qy 699 uGluTyrAlaGluThrIleLysHisSerLeuGlnLysIleLeuPheLeuLeuProPh 719  
Db 4581 TCAGTTGCTGAAAAAATGAAGGAG 4605  
Qy 719 eLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 735  
Db 4606 4652  
Qy 735 uPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerG1 755  
Db 4653 TCAGTTGCCAGGAGGAGGTTGAAGAAAGAACGGAACCTCTTCAGACTTTGCTCTCTGA 4712  
Qy 755 uLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuG1 775  
Db 4713 TGTGAGTGAGCTGTTAAAGACAAACTCAT 4748  
Qy 775 nHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLe 795  
Db 4749 GGAAGAGCTGCAGAGTTTGGAAAAAGACTCACAGGCATGCTTTTGACAAATGTGAGCT 4808  
Qy 795 uArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeuTh 815  
Db 4809 GGAATACCAATTCACAACTGAATAAGAGAAAGAAATGTTGTCAGGAATCTGAAG 4868  
Qy 815 rArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAl 835  
Db 4869 CCTGCAGCCAGACTG 4922  
Qy 835 aLeuGlyAlaLeu 851  
Db 4923 CTGGAGGCCGCACTGTTGGAGAAAGGTGAGTTGCAATTTGAGGCTGAGC 4973  
Qy 851 aAlaGlnAspIleValAspGlyHisArgGluLysThrLeuLeuTrpGlnLeu1 871  
Db 4974 AACACAGAGGAGAGTGCATCAGCTGAGA 5030  
Qy 871 eTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGlnLysTyrTrpAr 891  
Db 5031 TGAGCCCGATGAAAGAGACGCTGCACATCGCAGAGAACTGAAAGAACCGGAGCGGA 5090  
Qy 891 gArgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGluLeuMetArgAr 911  
Db 5091 GAATGATTCATTAAGGATAAAGTTGAGAACCTTGAAGAGGAATTTGCAGATGTCAGAGA 5150  
Qy 911 gHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrVa 931  
Db 5151 AAACAGAGGAGTGTGATTCCTGTGATGCC 5198  
Qy 931 lLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgAr 951  
Db 5199 AGAGACTCTAAAAACACAAATAGAGAGATGGCCAGAGACCTGAAAGTTTGAATTA 5256  
Qy 951 gTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleLeuThrIleGlnArgTr 971  
Db 5257 5277  
Qy 971 pTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnPheValGluLeuArgGluAl 991  
Db 5278 5333  
Qy 991 aAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAla1 1011

Db 5334 AGACAAGTTACTCTCTTCATTT 5372  
Qy 1011 aAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnLysAlaAlaIleSerTyrIleG1 1031  
Db 5373 AAAGGACAGACAGATACAGATCAAGAAATCTAAACTGCGTGGAGATGCTTCA 5432  
Qy 1031 nMetGlnTrpArgThrTyrGlnLeuGlyArgIle 1042  
Db 5433 GAATCAGTTAAAG 5486  
Qy 1043 5487  
Db 5487 AATTATAGAGCCACAGACAGAGCTAGACCCACCAATAGAGGAAGACATCAGCTG 5544  
Qy 1048 uArgGlnArgAspLeuIleMetPheValGlnArgArgMet 1064  
Db 5545 5597  
Qy 1064 sTrpSerMetLeuGluGlnArgLysGluPheGlnGln 1080  
Db 5598 GCTCTGTGCTTACACAACTGAAGGAAGTGAGCATCATCGATTTTACTTTAAGGGTAG 5657  
Qy 1080 aAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAs 1100  
Db 5658 AGTGAGAGAACCTTGAAGA 5705  
Qy 1100 pTyrLeuAlaLeuArgSerSerValLeuLys 1110  
Db 5706 GCATGCACTCTTGAGGAGAGAAATTCAAAGGAGAGGTAGAGACCCCTAAAGCAAAAT 5765  
Qy 1111 1127  
Db 5766 AGAAGGATGACCCCAAGCTCTGAGAGGCTCTGGAATTAGATCTTGTACTATAAGGTGAGA 5825  
Qy 1127 sTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLys 1147  
Db 5826 AAAAGAAAATCTGACAAATGAATTACAAAAAGACGAGGCAATATCTGAATTAGAAAT 5885  
Qy 1147 sMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLys 1167  
Db 5886 AATAAATTCATTCATTTGAAAATATTTTCAAGAAAAAGAG 5930  
Qy 1167 sArgTyrArgMetArgGln 1181  
Db 5931 CAAAGTACAGATGAAGAAAATCAAGCACTGCCATCGAGATGCTTCAACACAAATTAAG 5990  
Qy 1181 atyLeuArgThrArgLysCysIleIle 1200  
Db 5991 AGAGTCAATGAGAGAGTGGCAGCCCTGCATAATGACCAAGAGCCTGTAAGGCCAAGA 6050  
Qy 1200 uGln 1201  
Db 6051 GCAGAAATCTTAGTAGTCAAGTAGAGTGTCTTGAACCTTGAGAGGCTCAGTTGCTACAGG 6110  
Qy 1202 1220  
Db 6111 CCTTGATGAGCCAAAATAATATATATTTGTTTTCATCTTCAGTGAATGGCTCATTC 6170  
Qy 1220 nIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
Db 6171 AGAAGTGAAGATGGCAAGCAGAAAACCTGGAGAGAGGATGAAGAAATCAGTAGACTGAA 6230  
Qy 1240 sLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluAr 1260  
Db 6231 AAATCAAAATTCAGACCAAGAGAGAGCTGTCTCTTAAACTGTCCCAGGTGGAGGAGACA 6290  
Qy 1260 sGln 1276  
Db 6291 CCACTTGGAGAGAGCAAACTTAGAACTGAGAAATCTGACAGTGAATTTGGAGCAGAA 6350  
Qy 1276 sPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAl 1296  
Db 6351 G 6392



OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10136 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-2

Alignment Scores:  
 Pred. No.: 1,15e-13 Length: 10136  
 Score: 259.00 Matches: 374  
 Percent Similarity: 34.60% Conservative: 329  
 Best Local Similarity: 18.41% Mismatches: 791  
 Query Match: 2.72% Indels: 544  
 DB: 83  
 Gaps: 1

US-09-914-698-1 (1-1861) x US-08-353-700-2 (1-10136)

QY 105 HisProProGlnAlaProLeuValGluLysAsnValTyrLys----- 119  
 Db 4299 CACCCTGTGCTTGGCTCCATTGACGAGAGTAATCTCTACGACACTTGACATGTCA 4358  
 QY 120 -----ThrProGlnGluProValTyrLys----- 130  
 Db 4359 GACAAAGATTCAATGCATCTTCCGAAATTCGAAAGAAATCTTATCTTTACAAAGT 4418  
 QY 131 GlnProArgSerLeuLysGluAsnLeuSerProMetThrPro-----GlyAsnLeuLeu 148  
 Db 4419 GAACACAAAATTTTACATGATCAGCACTGTCTCAGATGAGCTCTAAATGTCTCAGAGCTGCGAG 4478  
 QY 149 AspValIleAspAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAla 168  
 Db 4479 ACCTATGTTGACTCATTA-----AAGCCCGAAAT 4508  
 QY 169 ThrIlePheProAspAsnLeuAlaIleThrProThrProThrLeuLys----- 184  
 Db 4509 TTGGCTTGTTCACCAATCTGAGAAATTTCAAGGTGACTTGTGAAAGGAGATGCGAGCTG 4568  
 QY 185 -----GlyAsnValLysSerCysAlaAsnAspMetArgProArgIleThr 200  
 Db 4569 GCGCTGGAGGAGGGCTCGTTCATCCCTGCTATCCTCTTGTGCTGCTCAGACTCTAGT 4628  
 QY 201 ProAspAspLeuGluAsp-----GlnProAlaThrAsnLysThrPheAspVal 216  
 Db 4629 CTTAGCAGTTGGGAGGACTCTCTCTTTACAGAGCTCTTTAGAACAGACAGAGATATG 4688  
 QY 217 LysHisSerGluThrIle-----AsnIleSerLeuAspThrLeu 229  
 Db 4689 TCTCTTTTGAGTAATTTAGAGGGGCTGTTTCAGAAACCCAGTGCAGTGTAGATGAAGTA 4748  
 QY 230 AspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal 249  
 Db 4749 TTTTGCAGCAGTCTG-----CAGACCTATGTTGACTCATTAAGCCGGAATTTGGTC 4802  
 QY 250 HisAlaThrHisThrArgAlaLeuAla-----CysIleHisGlu----- 262

Db 4803 TTGTCAACGAATCTGAGAAACTTTTCAAGGTGACTTGGTGAAGGAGATGCGAGCTGGGCTTG 4862  
 QY 263 GluGluGlyProSerPro-----ProArgThrProThrLysSer 275  
 Db 4863 GAGGAGGGGCTCGTTCCATCCCTGTGCTCTCTGTGCTGACAGCTCTAGTCTTAGC 4922  
 QY 276 AlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTyrSer 295  
 Db 4923 AGTTTGGGAGAC-----TCCTCCTTTTACAGAGCTCTTTTA 4958  
 QY 296 GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySer 315  
 Db 4959 GAACAGACAGAGATATGCTCTCTTTTTCAGT-----AATTAGAGGGGTGTTTTCAGCAAA 5015  
 QY 316 MetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlu 335  
 Db 5016 CAGTGCAGTGTAGATGAAGTATTTTTCAGCAGCTCTGCGAG----- 5057  
 QY 336 GlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeu 355  
 Db 5058 -----GAGAACTGACACGAGAAAGAAACCCCTTCGCGCCCGCAGCG 5096  
 QY 356 AlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe 375  
 Db 5097 AAGGTGTGTTGAAGAGCTT-----GAGTCCCTCTGTGAGGTGTACCGGAG 5141  
 QY 376 AsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHisLys 395  
 Db 5142 TCCCTCGAAGCTAGAGAGAAATGGAAGTCAAGGATATGAAATAAGGAAATT 5201  
 QY 396 ArgArgSerHisGluLeuSerPheSer-----AspAlaProSerAsnGlu 410  
 Db 5202 CAAGAGCTCGAGCAGTATTAAAGTTCTGAAAGGCAAGAGCTTGACTGCTTAGGAAGCAG 5261  
 QY 411 SerLeuTyrArgAsnGluThrValAlaIleSerProProLys----- 424  
 Db 5262 TATTGTGCAAAAATGA-ACAGTGGCAACAGAAAGTGCACGCGTACTCTGGAATGA 5320  
 QY 425 -----LysGlnArgValGluAspThrThrLeuPro-----ArgSer 436  
 Db 5321 GTCCAAGTTGGCGGAGAAAGAAACAGACGGAACACTGTCTGAGTGGAGTAGC 5380  
 QY 437 AlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaIleProHisAlaGln 456  
 Db 5381 ACGACTCCAGCTACAAAGTCTGGACTTAAGTCTCGGTCTTCTGCTGGCA-TCGACACAG 5439  
 QY 457 SerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArg 476  
 Db 5440 AAGATGCTATTTCAGGCGCAATAGAGCTGTGAC-ATATCAAAAGAACATATCTTCAGAA 5498  
 QY 477 LysValArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMet 496  
 Db 5499 ACTCAGAAAGAACCAACCAAGCATGATGTTTCATCAGATTGTGTATAAAGATGCTCAGCAG 5558  
 QY 497 GlnThrCysIleAspProAspProPheAlaAlaThrThrThrIleAspProPhe----- 514  
 Db 5559 GACCTCAATCTAGACATTGAGAAATAACTGAGACTGTGTGCTGAGTCAAAACCCAGAGAG 5618  
 QY 515 -----LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529  
 Db 5619 TGCTCTGGGGAACAGTCCCGCAGATACCAATTAT---GAGCCTCCAGGGGGAAGATAAAC 5675  
 QY 530 GlnAlaAspPheLys-----LysTrpLeuAsnAlaLeuValSer 542  
 Db 5676 CAGGGCTCTTCGAATGCGATTCTGAAATGCTCATTTCTGGTCTCTTAATGCTTTGTA--- 5732  
 QY 543 IleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLysLysPheAsn 562  
 Db 5733 ---CCTATGCAATTC-----CTGGGGAATCAGGAAGATATCCATATCTTCAACTG 5780  
 QY 563 GluVal----- 564  
 Db 5781 CCGGTTAAAGAGACATCAAAATGAGAAATTTGAGATTACTTCTGTGATGAGGACCGTGC 5840









5440 AAGATGCTATTACAGCGCCGAAATGAGAGCTGTGAC-ATATCAAAAGAAACATATCTTCAGAA 5498  
 QY  
 477 LysValArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMet 496  
 Db  
 5499 ACTACAGAAAGACCAACCAAGCATGATGTTTCATCAGATTTGTGATAAAGATGCTCAGCAG 5558  
 QY  
 497 GlnThrCysIleAsnProAspPheAlaAlaThrThrIleAspProphe----- 514  
 Db  
 5559 GACCTCAATCAGACATTGAGAAATAACTGAGACTGGTGCAGTGAACCCACAGGAGAG 5618  
 QY  
 515 -----LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529  
 Db  
 5619 TGCTCTGGGGAACAGTCCCGCATACCAATTAT---GAGCCTCCAGGGGAAGATAAACC 5675  
 QY  
 530 GlnAlaAspPheLys-----LysTyrLeuAsnAlaLeuValSer 542  
 Db  
 5676 CAGGCTCTTCAGAAATGATTTCTGAATTGTCATTTCTGGTCTTAATGCTTTGGTA--- 5732  
 QY  
 543 IleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsn 562  
 Db  
 5733 ---CCTATGGATTTC-----CTGGGAATCAGGAAGATATCCATAATCTTCAACTG 5780  
 QY  
 563 GluVal----- 564  
 Db  
 5781 CGGGTAAAGAGACATCAAAATGAGAAATTTGAGATTACTTATGTCATGATAGAGCGCGTGAC 5840  
 QY  
 565 ArgAsnLysGluLeuValValAlaProThrLysGluGlnGlnSerMetAsnTyrIleuThr 584  
 Db  
 5841 AGAAAGTTGAAAGTTTCTTAATGAATGAAGAAATTAGACTCAAACTCCATTTACAG 5900  
 QY  
 595 LysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu----- 598  
 Db  
 5901 GAGGTACAACTAATGACCAAAATTAAGCATGATGAGAAATTTGGAATAATAGTTGGGAA 5960  
 QY  
 599 -----PhePheSerGluGlnMetArg---LeuProCysSerLysVal 611  
 Db  
 5961 CTTAAGAAAGAAACTCAGATTTAAGTGAATAATTTGTAATATTTTCTGTGATCACCAG 6020  
 QY  
 612 AlaValTyrValAsnLysGlnAlaLeuArg---IleArgSerAspArgAsnLeuHisLeu 630  
 Db  
 6021 GAGTTACTCCAGAGATGAGAACTTCTGAAGGCCTCAATCTGATTTAGAATATGATGCA 6080  
 QY  
 631 AspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTripleu 650  
 Db  
 6081 GATAAATCATCAGCTGAAGATATTCGAGATATGTCGACCAAGGTGAATCAGACGTGGAG 6140  
 QY  
 651 ArgLeuGlyLeuGluVal-----Val 657  
 Db  
 6141 GAGAGATTCTTGTGATGTGAAATAGAGCTGAGTATGATCGGAGAAAGCTAGCAATT 6200  
 QY  
 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db  
 6201 GAGCATGAAGCCCTCTACCTGGAGCTGACTTAGGAGGTATGTTCAACAGAGAAGCTATGT 6260  
 QY  
 678 LeuAsnArgLeuPheArgAsnLys-----CysGluGluGlnArg----- 690  
 Db  
 6261 TTAGAAAAGAGCAATGAAATAAGCAGAGAGTTATGTCGCTCGCTTGAAGAAGACTCTCA 6320  
 QY  
 691 -----TyrSerLysAlaTyr 695  
 Db  
 6321 GTGCTCAAGTGAAGAAACCCAGCTTCGTGGAGAAATTTAGATACTATGTCAAAAAACC 6380  
 QY  
 696 ThrLeuThrGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPhe 715  
 Db  
 6381 ACGGACTGGATCAGTTCTGTGAAATAATGAGGAG----- 6416  
 QY  
 716 LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro----- 733  
 Db  
 6417 -----AAACACAAGAGCTTGTGATCTCATCAAGTACAGTGT 6452  
 QY  
 734 -----CysLeuPheValLysSerProHisLysGluThrLysAspIleLeuLeuArg 751  
 Db  
 6453 CTCATTGCAATTCAGTGGCAGAGCGAGAGTGAAGGAAAGACGGAACCTCTTCAGACT 6512

QY 752 PheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGly 771  
 Db 6513 TTGTCTCTTGATGTGAGTGAAGTGTAAAAGACAAAACACTCAT----- 6554  
 QY 772 TyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeu 791  
 Db 6555 -----CTCCAGGAAAGCTGCAGAGTTTGGAAAGGACTCACGCACTGTCCTTTGACA 6608  
 QY 792 AlaValAspLeuArgAspGlyValArgLeuThrArgValGluValIleLeuLeuArg 811  
 Db 6609 AAATGTGAGCTGGAACCAACCAATTCACAACTGAATAAAGAGAAAGAAATGCTTGTCAAG 6668  
 QY 812 AspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsn 831  
 Db 6669 GAATCTGAAGCCCTGCAGCCAGACTG---AGTGAATCAGATTATGAAAAGCTG---AAT 6722  
 QY 832 VallysLeuAlaLeuGlyAlaLeu-----GlyGluAlaAsnPheGlnLeuGly 847  
 Db 6723 CTCTCCAGGCTTGGAGCGCCACTGCTGGAGAAAGGTGAGTTGCGCATTTGAGGCTGAGC 6782  
 QY 848 GlyAspIleAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeu 867  
 Db 6783 -----TCAACACAGGAGGAAAGTGCATCAGCTGAGA---AGAGGCATCGAGAACTG 6830  
 QY 868 TrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGln 887  
 Db 6831 AGAGTTCGATGAGCGCGATGAAAGAGCAGCTGCATCGCAGAGAAATGAAAGAA 6890  
 QY 888 LysTrpTrpArgHisTrpLeuHisValIleGlnArgArgIleArgHisLysGlu 907  
 Db 6891 CGCGAGCGGAGATGATTCATTAAGATTAAGTTGGAACCTTGAAGGGAATTTGAGC 6950  
 QY 908 LeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMet 927  
 Db 6951 ATGTCAAGAAACACAGAGGAGTGTGATTTCTGATGCC-----GAGAATTC 6998  
 QY 928 ArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947  
 Db 6999 AAACAGAGTAGAGACTCTAAACACAAATAAGAGAGATGGCCAGAGCCTGAAGATT 7058  
 QY 948 PheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThr 967  
 Db 7059 TTTGAAATTA-----GACCTTGTCTACG 7079  
 QY 968 IleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGlu 987  
 Db 7080 TTAAGGTCT-----GAAAAAGAAAAATCTGACAAAAACAAATACAGAAAAACAAGGTCTCAG 7133  
 QY 988 LeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLys 1007  
 Db 7134 TTGTCAGAACTAGACAAAGTTACTCTCTTCATTT-----AAAAAGT 7172  
 QY 1008 LeuLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAla 1027  
 Db 7173 CTGTTAGAGAAAGGAGGAGCAGAGATCAGATCAAGAGAGATCTAAACCTGCATG 7232  
 QY 1028 SerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIle----- 1042  
 Db 7233 GAGATGCTTCAGATCAGTTAAAG-----GAGCTAAATGAGGCAGTAGCAGCCTTGTGT 7286  
 QY 1043 -----GlnArg 1044  
 Db 7287 GGTGACCAAGAAATTAAGAGCCGACAGAACAGAGCTTAGACCCACCAATAGAGGAGAG 7346  
 QY 1045 HisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMet----- 1061  
 Db 7347 CATCAGCTG-----AGAAATAGCATTAAGAGCTGAGAGCCCGCTAGAGAGCTGAT 7397  
 QY 1062 ---ArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGln----- 1076  
 Db 7398 GAAAAGAGCAGCTCTGCTCTTACAACTGAAGGAAAGTGAAGCATCATGTCAGATTTA 7457

QY	1077	LeuLysArgAlaalaalleasnlleGlnArgTyrArgAlaLysLeuSerMetArgLys	1095
Db	7458	CTTAAAGGGTAGCTGGAGAACCTTTGAAAGA-----GAGCTAGAGATAGCCAGG	7505
QY	1097	CysAsnAlaaspTyrLeuAlaLeuArgSerSerValLeuLys-----	1110
Db	7506	ACAAACAAGAGCTGCAGCTCTTGAGCGAGAGAAATCCAAAGAGAGGTTAGAGACCTTA	7565
QY	1111	-----ValGlnAlaTyrArgLysAlaThrIleGlnMetArgIle	1123
Db	7566	AAAGCAAAAATAGAGGGATGACCCAAAGCTCGAGAGGCTCGAATATAGATGTTGTACT	7625
QY	1124	AspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnArgLeuArg	1143
Db	7626	ATAAGGTCCGAAAAAGAAATCTGCACAAATGAATTTACAAAAGCAGCAGGCAATATCT	7685
QY	1144	AlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIle	1163
Db	7686	GAATTAGAAATAATAAATTCATCATTTGAAATAATTTGCAAGAAAAAGAG-----	7735
QY	1164	LeuValGlnLysArgTyrArgMetArgGln-----GlnMetIleGln	1177
Db	7737	-----CAAGAGAAAGTACAGATGAAGAAAAATCAAGCACTGCCATGGAGATGCTTCAA	7790
QY	1178	AspArgAsnAlaTyrLeuArgThrArgLysCysIleIle-----AsnValGlnArgArgTyr	1196
Db	7791	ACACAATTAAGAGCTCAATGAGAGAGCTGGCAGCCCTGCATATATGCCAAGAGCCTGT	7850
QY	1197	ArgAlaThrLeuGln-----	1201
Db	7851	AAGGCCAAAGAGCAGCAATCTTAGTAGCTAGAGTAGAGTGCTTGCAACTTGAGAGGCTCAG	7910
QY	1202	-----MetArgArgGluArgLysAsnTyrLeuHiIleuGlnThrThrThrLys	1217
Db	7911	TTGCTTCAAGGCCTTGATGAGGCCAAATAATATATATTTTGTTCATCTTCAGTGAAA	7970
QY	1218	---ArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPhe	1236
Db	7971	GGCCTCATTCAGAAAGTAGAAGATGGCAAGCAGAAACTGGAGAGAAGAGGATGAAGAAATC	8030
QY	1237	LeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMet	1256
Db	8031	AGTAGCTGAAAAATCAAATTTCAAGACCAAGACAGCGCTTGCTCTAAACTGTCTCCAGGTG	8090
QY	1257	ArgLysGluArgGln-----GlnTyrLeuHiIleuArgGluValThrIleLys	1272
Db	8091	GAAGGAGAGCACCAACTTTTGAAGGACCAAACTTAGAACTGAGAAATCTGACGGTGGAA	8150
QY	1273	LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly	1292
Db	8151	TTGAGACAGAG-----ATCCAGTGTCTACAATCCAAAAATGCCTCT	8192
QY	1293	ThrGlnAlaAlaValSerCysLeuGlnMetHisTyrArgAsn-----HisLeu	1308
Db	8193	TTGCAGACACATAGAAAGTCTGCAGAGTTCTTACARGAATCTAGAGAATGAGCTTGAA	8252
QY	1309	LeuArgLysArgGluArgAsnSerPheLeu-----	1318
Db	8253	TTGACAAAAATGGACAAAATGTCTTTGTTGAAAAAGGTAAACAAAATGACTGCAAGAGAA	8312
QY	1319	-----GlnLeuArgGlnAlaAlaIleThrLeuGlnArgArg	1330
Db	8313	ACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAACACAGCAGAGTGCAGAAAGAA	8372
QY	1331	TyrArgAlaArgLeuAsn-----MetIleLysGlnLeuLys	1342
Db	8373	CTCAGTCGAGAGAAAATAGCTAGCTGGAGAGTTGCAGTTACTGTTGGAGAAATAAAG	8432
QY	1343	SerTyrAla---GlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLys	1361
Db	8433	AGCAGCAAGATCAATTTGAAGGAGCTCACATAGAAAATAGTGAATTTGAAGAGAGCCTA	8492
QY	1362	LvsAlaMetGlnLys---GlnValValLeuTyrGlnLysGlnArgGluAlaIleLys	1380

[illegible]



Db 1348 CTCGGTTACCACTCCGTTACCAACCGCGCTCCACCTCCAGCTCAAGCTGATGCAGATG 1289  
Qy 1139 GlnGlnArgLeuArgAlaIleMetLysMetArgGlnArgGluAsnTyrLeuArg--- 1157  
Db 1288 CAGATG---CTGAAGCAGATCCTGAGCTTTCAATAGACCTGCAAGATCCTTTAAGACAT 1232  
Qy 1158 -----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMet 1175  
Db 1231 CCTCAATCGTCAATTGAGATTGTGGCTTGCTACTGCTGATGCTTTAGCAGAGGCTT 1172  
Qy 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle----- 1190  
Db 1171 TTGCAGAGGCTCCATTACTCTGTGAAGAGGTCTTTAATGCTCTCTCAATGCGCTCCCTTATAC 1112  
Qy 1191 -----AsnValGlnArgArgTyrArgAlaThrLeuGlnMet 1202  
Db 1111 CACTCATTGCTAGAGCTGAACAGATGACAGCGGCTGCGCGGCGGCGGCGAC 1052  
Qy 1203 ArgArgGluArg-----LysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
Db 1051 CAGCGCTCGCGCTCTCTCTCCACACCTCCACCTCTCGGCCACACCGAGCGCTG--- 995  
Qy 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu-----Phe 1236  
Db 994 -----CGCGCGAGCTGCGAGCGGCGGCGAGCGAGCGAGCTGCGGCGCTC 950  
Qy 1237 LeuGlnLeuLysLysValThrLeuValGlnLysArgArg-----AlaLeuLeu 1254  
Db 949 CTCTCTCTCAGCGGTCTACTCTCTCTGTCACCGTAGAACCAACACACCGAGTCTCTC 890  
Qy 1255 GlnMetArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrLysLeuGln 1274  
Db 889 CAGCTCGGCGAGCGGCGGCGGCGAGCGAGCGGCGAGCAGATCCTCCATATCTC 830  
Qy 1275 ArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGln 1294  
Db 829 CGA-----GGCCACCAAGACCTCCGAGGCTCCAACTGCCACCGAGTCTCTCAA 779  
Qy 1295 AlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArg 1314  
Db 778 GTCCGCCACCGAGTCTCTCAAGTCCACCTCGC-----CTCTGCGAGCGGCGAGCGG 725  
Qy 1315 AsnSerPheLeuGlnLeuArgGln-----AlaAlaIleThrLeuGln 1328  
Db 724 -----CTGGGCTGCGAGCAGCAGCTCCTCCAGCTACCTGCGCTCTCTCTGCGAG 674  
Qy 1329 ArgArgTyrArgAlaArgLeu-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
Db 673 CGCAGCAGCGGTGCGGTGCGGCGAGCAGCTCCACCACTCCACCCCGCGGCGGCGAG 614  
Qy 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
Db 613 CGGCGAG----- 608  
Qy 1368 ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGly 1387  
Db 607 -----CTGCAGCAGCAGCAGCAGC-----CTCCAGCTCCACCGCTCCACCGAGG 560  
Qy 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
Db 559 CAA-----GAGCTGCACCAAGATCATCGAAT---CTAAATCAA 524  
Qy 1408 LeuGlnLysTyrTrpArgSerIleArg---AspMetArgLeuCysLysAlaGlyTyrArg 1426  
Db 523 ATAAATCGTCTCCACCGCCCAAGTCCGCCAAGACTCCAACTC-----CGC 479  
Qy 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArg 1446  
Db 478 CAACTCCACTCCGAGACCTC-----CGGCCAATCCAAAGTCCGA 440  
Qy 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466  
Db 439 GTCCAGCACTGCTCTCTGCTC-----CTGGCGCTGCTCCAGCAGCAGCAGC 395

Qy 1467 ArgAlaThrLeuLeuMetArgGlnGln-----ArgArgGluPheGluMetLysArg 1483  
Db 394 CAGCACCCTCTCGCGCACCAACACAGCACCACCACTCCCGCGCGCCAGCTCCCATCCACCGT 335  
Qy 1484 ArgAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db 334 CATCATCCAGTCCCATTCATCATCGTCATCGTGTGCT----- 296  
Qy 1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
Db 295 -----CATCATAGTCATATTCCTCAATCTCCACCGCGCGCGCGTCCGC 251  
Qy 1524 ArgSerMetLysGlnAlaArgGlnPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 250 CGTCGGCTCTCCAGCACCGCCAG-----CTC 224  
Qy 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGlnArgAsnCysPheGlnLeuLeu 1563  
Db 223 CCCAGGCGCTCCGCCCTCCGACACCGTAAGAC-----CTG 185  
Qy 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 184 CACTGCTCCATGCCGA-ACGCAACACCATGCCCATGGATCATATAAACCATGCCAGGCATA 126  
Qy 1584 GlnAlaLeuMetThrPro 1589  
Db 125 CAGAAAGTCTTAGTCT 108  
RESULT 12  
US-08-864-038A-2/c  
; Sequence 2, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS PRODUCING SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; STREET: Isshinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: P-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3331  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA to mRNA

895	LeuHisValValIleGlnArgArgIleArgHisLys-----GluLeuMetArgHis	912
QY		
2192	ATCATATTGTTTCACCGCCGCTGATCGCGACCGCTGAACACACAGATCCGCGCCAC	2133
DB		
913	ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys	932
QY		
2132	CGC-----CTGATCCACAGATCCACCGCACCTCCAGATTCGCGCACCGGATCCAC	2082
DB		
933	LeuPheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyr	952
QY		
2081	CTCCGCGGATCCGCCACACACAGATCGGCCAOCCTCCATTTCGTTCCACCAACCGATC	2022
DB		
953	LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTyrTrp	972
QY		
2021	CATTATCGAAATCAGTTCATCAGCACCGCTCAC-----CGTACATAGCTG	1977
DB		
973	ArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAla	992
QY		
1976	CGGCTCAGCAGCGGCGAGCGGCAACATCAGCAGCATCTCCAGCAGCAGCAGCAGCGG	1917
DB		
993	IlePheLeuGln-----ArgIleTrpArgArgArgLeuPheAlaLysLys	1007
QY		
1916	CGCGCGCTCGAGCTACAGCACTAGCTCCGTTACCGTCGCCATCGCTCCACGCTCTCTAC	1857
DB		
1008	LeuLeuAlaAlaAlaGluThrAlaArgLeuGlnArg-----SerGlnLysGlnGlnAla	1025
QY		
1856	CTCTACTCTTCGTCCACCTCCACCAAAATCCAGCGCGGCGGTGCAGCAGCGGCGAGCTG	1797
DB		
1026	AlaAlaSer-----	1028
QY		
1796	CTGCTGATGATCCACCACCAATCCTCTCCGAATCCTCCACCGCAGACCTACTCCGAATC	1737
DB		
1029	-----TyrIleGlnMetGlnTrpArgThrTyrGln	1038
QY		
1736	CTCCACCCATTCACCTCCCATCCACCTCCAGCAGCAGCTGCAGCAGCAGCAGCAGCG	1677
DB		
1039	LeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeu--IleMetPheVal	1057
QY		
1676	CAGCAGCGGATCCACTCCAGCATTTGTCTCTCAAAGCCCTCTAAGTGTCTTCTCTC	1617
DB		
1058	GlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeu	1077
QY		
1616	CACCAC-----CTGATGCAGCAGCGGCGGCTGCAGCAGCTG	1578
DB		
1078	Lys-----ArgAlaAlaIleAsnIleGlnGlnArg-----	1087
QY		
1577	CGGCAGATCCCAACAGACCGCTAGTCTCTCAAGTCCACCAACCTCCTCCACCTG	1518
DB		
1088	-----TrpArgAlaLysLeuSerMetArg-----LysCysAsn	1098
QY		
1517	CACCAGCAGCAGTGTAGTCAGCGGCTTAAGCTCCGCTCCGCTCCACCTCCAGTCCAC	1458
DB		
1099	AlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTyrArgLysAlaThr	1118
QY		
1457	CTCCGCGCTCTGCAGCAGCAGCAGCAGTGTAGTCTCAGTCCAGCGCGCTCTCTC	1398
DB		





QY 1099 AlaAspTyrLeuAlaLeuArgSerSerValIleuLysValGlnAlaTyrArgLysAlaThr 1118  
Db 1457 CTCGGCTCTCCGACGACGAGCAGCAGTAGAGCTCCAGCTCCACCGCGCTCCTC 1398  
QY 1119 IleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeu 1138  
Db 1397 CTCGGTTACCACTCCGTTACCAACCGCTCCACCTCCAGCTGAAGCTGATGAGATG 1338  
QY 1139 GlnGlnArgLeuAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArg 1157  
Db 1337 CAGATG---CTGAAGCAGATGCTGAGCTTTTCAATAGACCTGCAAGATCTTTAAGACAT 1281  
QY 1158 -----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnMet 1175  
Db 1280 CCTCAATCTCAATTGAGATTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221  
QY 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle----- 1190  
Db 1220 TTGCAGAGCTCCATCTCTGAGAAGTCTTTAATGCTCTCTAATGCTCCCTTATATAC 1161  
QY 1191 -----AsnValGlnArgArgTyrArgAlaThrLeuGlnMet 1202  
Db 1160 CACTCATTTGCTAGAGCTGAAGCAGATGAGCAGCGCTGCGAGCGCAGCGCAGCAGC 1101  
QY 1203 ArgArgGluArg-----LysAsnTyrLeuHisLeuGlnThrThrTyrLysArgIleGln 1220  
Db 1100 CAGCGCTCCGCTCTCTCTCCACCACTCTCCACCTCTCCGCGCACCAACAGCGGTG--- 1044  
QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu-----Phe 1236  
Db 1043 -----CGCGCGAGCTGCGAGCGCGCAGCAGCAGCAGCTGTCGCGCTC 999  
QY 1237 LeuGlnLeuLysLysValThrLeuValValGlnLysArgArg-----AlaLeuLeu 1254  
Db 998 CTCTTCTCGACCGGTCTACCTCTCTCTCCACCGCTAGAAACCAACACCGAGTCTCCTC 939  
QY 1255 GlnMetArgLysGluArgGlnLysLeuHisLeuArgGluValThrIleLysLeuGln 1274  
Db 938 CACTCCGCGCAGCGCGCGCAGCAGCAGCGCAGCAGCAGCTCTCCATATCTC 879  
QY 1275 ArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGln 1294  
Db 878 CGA-----GGCCACCAAGCTCCGAGCGCTCCAGTCCGCGCAGCGCTCTCCAA 828  
QY 1295 AlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArg 1314  
Db 827 GTCCGCCACCGAGTCTCCAGTCCACCTGCGC-----CTCTGCGCGCGCAGCAGCGG 774  
QY 1315 AsnSerPheLeuGlnLeuArgGln-----AlaAlaIleThrLeuGln 1328  
Db 773 -----CTGCGGCTGCGAGCAGCTCTCCAGTCTACCTCGGCTCTCTCTCGAG 723  
QY 1329 ArgArgTyrArgAlaArgLeu-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
Db 722 CGGCGAGCGGCTCGGCTCGGCGAGCAGCTCCACCACTCCACCCCGCGCGCAGCAG 663  
QY 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
Db 662 CGGCAG----- 657  
QY 1368 ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArgGly 1387  
Db 656 -----CTGCAGCAGCAGCAGCAGC-----CTCCAGCTCCACGCTCCACCGAGG 609  
QY 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
Db 608 CAA-----GAGCTGCACCAAGATCATCGAAT---CTAATCAA 573  
QY 1408 LeuGlnLysTyrTyrArgSerIleArg-----AspMetArgLeuCysLysAlaGlyTyrArg 1426  
Db 572 ATAAATCGTCTCCACCGCAAGTCCGCGCAAGACCTCCCAAGTC-----CGC 528

QY 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArg 1446  
Db 527 CAAGTCCACCTCCGAGACCTC-----CGCCCAATCCAGTCCGA 489  
QY 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466  
Db 488 GTCCAGCACCTGCTCTCTGCTC-----CTGCGCTGCTCCAGCAGCAGCAGC 444  
QY 1467 ArgAlaThrLeuLeuMetArgGlnGln-----ArgArgGluPheGluMetLysArg 1483  
Db 443 CAGCACCACTCCCGCAGCAGCAGCAGCAGCAGCAGCTCCGCGCGCAGCTCCATTCACCGT 384  
QY 1484 ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db 383 CATCATCCAGTCCCATTCATCATCGTCTCTCT----- 345  
QY 1504 GlnAspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
Db 344 -----CATCATAGTCATATCCCAATCTCCACCGCGCGCGCTCGC 300  
QY 1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 299 CGTGGCTCTCCACCGACCGCAG-----CTC 273  
QY 1544 GlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db 272 CCAGGCGCTCCGCGCTCCGACCGTAAGAC-----CTG 234  
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 233 CACTGCTCATGCGGAA-ACGCAACCATGCTCATGATGATATAAACCATGCGCAGGCATA 175  
QY 1584 GlnAlaLeuMetThrPro 1589  
Db 174 CAGAAAGTCTTAGTCTCT 157  
RESULT 14  
US-08-956-171E-63  
; Sequence 63, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hyman  
; REGISTRATION NUMBER: 46,789



QY 568 -----GluLeuValValAlaProThrLys 575  
 DB 3605 TGAACAAAGATGGTGGCAACAAACAAAGTAACTCAAGATGATTAATAGTCAATACACAA 3664  
 QY 576 GluGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAla 595  
 DB 3665 CAGCATCAATTGCAGAT-----ACATAAACTTAACAAC 3700  
 QY 596 ValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrVal 615  
 DB 3701 AAGCAGACTATTATTAATCAAGATGCAATCATCTTAAACTGCAATCGTGCATCTC 3760  
 QY 616 AsnLysGlnAlaLeuArgIleArgSerAspAsnLeuHisLeuAspValValMetGln 635  
 DB 3761 AAGCGATATTGATGGTTAGTAATAATTACAAAGCTGCATTAATGATTAATCAAGCAG 3820  
 QY 636 ArgThrIleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuGlu 655  
 DB 3821 CAAATTGCTGAATTAGATA-----CTAAAGCTCAAGAA-AAAG 3855  
 QY 656 ValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThr 675  
 DB 3856 GTTACAGCAGCACAAACAAAGTAAAGATGACCAAGATGAAGTTGCCAGCACTTGTAACT 3915  
 QY 676 PheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyr 695  
 DB 3916 AAAATTAAACATGATAAAATAATGCAATCCGAGAAATTAATAAAACAACTACACACAA 3975  
 QY 696 ThrLeuThrGluGlu-----TyrAlaGlu 703  
 DB 3976 GGTGTCAACCTGAAAGAAATAGTATCGTATCGAGTGTAGAACAGATCTGATTACACCA 4035  
 QY 704 ThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAla 723  
 DB 4036 ACAGTTAAACCTCAAGCGCAACAAAGATATATCCAGCAGTT-----ACAACT 4083  
 QY 724 LysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerProHisLys 743  
 DB 4084 CTTAAACCAACAAATTAAGTCAATATGATCATTA-----CAA 4122  
 QY 744 GluThrLysAspIleLeuArgPheSerSerGluLeuLeuAlaAsn-IleGlyAspIle 763  
 DB 4123 GATGAAAGATGTA-----GCAATGATTAATAATTGGT 4155  
 QY 763 eThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGln 783  
 DB 4156 AAAATTGAACAAAGGCAATTAAGATATGATGCGCAACAAACAA----- 4201  
 QY 783 uPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrAr 803  
 DB 4202 -----ATGCACAAAGTAGAGCCCATTAACAA 4227  
 QY 803 gValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgValProAlaIle 823  
 DB 4228 AAAGCAATCAATGATTAATCAACTACCTGCTACACAGCTTAAGCAGCAGCTCTT 4287  
 QY 823 eSerArgLeuGlnArgIlePheAsnValLysLeu----- 834  
 DB 4288 GAAGAAATTTGACCAAGATGTTCTCAAGCACAAATTTGATCAAGCACCTTTAAATCCTGATACA 4347  
 QY 835 -----AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGln 847  
 DB 4348 ACAATGAAAGATGACGGAAGCTATTGAACCTATTAAATGACAGTAAAGTTTCTGCTGTT 4407  
 QY 847 yGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLe 867  
 DB 4408 AAAGCAATTGAAGCGCACACAGCTG-----CACAG----- 4438  
 QY 867 uTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGln 887  
 DB 4439 -----ATTGAAAGAGTATAAAACGAAAGAAATCTCAAAATTTGAAAAATATTACT 4488

QY 887 nLysTrpTrpArgArgHisTyrLeuHisValValIleGlnArgArgIleArgHisLysGln 907  
 DB 4489 GACTCTACGCAACAAATAATGATGCTTATATGAGTAAAGTAAACAGCTGCACACACTAGA 4548  
 QY 907 uLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMe 927  
 DB 4549 AAAGCTCAAA-----ATGTACAG-----TTTCAATGCAACAAAT 4584  
 QY 927 tArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947  
 DB 4585 GAAGAAGTAGCAGAAAGCTGATGCAGCAG-----TAGATCAGCT 4623  
 QY 947 sPheThrArgArgTyrLeuAlaGlnLysGlnLysTyrGlnSerTyrHisSerIleIleTh 967  
 DB 4624 CAAAGCAAGTTTACATGACATCCAGTTGTTAAATCAAAACAGGAAGTTGCTGATACA 4683  
 QY 967 rIleGlnArgTrp---TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVa 986  
 DB 4684 AAATCAAAAGTATTAGATAAAATCAATGATCAATCAACACAGCAAAAGTTAAACCTGCA 4743  
 QY 986 lGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLys 1006  
 DB 4744 GCTGATACGGAAGTAGAAAACGCAATATAATACACGTAAACAGAAATTCAAAATAGCAAT 4803  
 QY 1006 sLysLeuLeuAlaAlaAlaGluThrAlaArgLeuGln-----ArgSerGlnLys 1022  
 DB 4804 GCTTCAACTACAGAAAGAAAACAGCTGCATATACAGAAATTAGATACATAAAAGCAAGAA 4863  
 QY 1022 sGlnGlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIle 1042  
 DB 4864 GCAAGAA-----CAAACTCTTGATGCTGCAAAATACAAACAGTGTATCAACACAGCT 4914  
 QY 1042 eGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe-----ValGln 1058  
 DB 4915 AAAGCAATAGTATTGCTGCA-----TTAATCAAGTTTCAAGCTGCCCAACCTAAG 4965  
 QY 1058 nArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLys 1078  
 DB 4966 AAATCGATGCAAGCGGAATCGCTCAAAAGACAGTGAACGTAACACAGCAATTTGAA 5025  
 QY 1078 sArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLysCysAs 1098  
 DB 5026 GCAATGAATGATTCGACTACTGAAAGCAACAAACAGCAGCGA----- 5065  
 QY 1098 nAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTyrArgLysAlaTh 1118  
 DB 5066 -----AAACAAAGTGGATCAAG---CAGTAGTTTACTGCAACACGCT 5103  
 QY 1118 rIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLe 1138  
 DB 5104 GATATAGATATGCTGCAGCAACAAATGATGTGATATATGCNAAACTTACAAATGAAGCT 5163  
 QY 1138 uGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLys 1158  
 DB 5164 ACAATCGCAGCATTTACACCTGATGCAAAATGTTAAACCCAGCAGCAAAACAAAG----- 5215  
 QY 1158 uArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAs 1178  
 DB 5216 -----CAATTGCAATAAAGTACAAAGCTCAAGCTCAAGAAACAGCAATTTGATGGA 5259  
 QY 1178 pArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTyrArgAl 1198  
 DB 5260 AATAAGCGCTCAACACTGAAGAAAHACAGCTGCTAAACCAACAAAGTTCAAACTGAAAAA 5319  
 QY 1198 aThr-----LeuGlnMetArgArg-----Gln 1205  
 DB 5320 ACAACAGCTGATCGCAATAGATGCAGCACATACAAATGCGGAAGTTGAAGCGGCTAAA 5379  
 QY 1205 uArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGlnIleLysPheArgAl 1225  
 DB 5380 AAAGCAGCAATTTGCTTAAATTTGAAGCGATTTCAGCCAGCAACAACTAAAGATAATGCG 5439  
 QY 1225 aLysArgGluMetLysLysGlnArgAlaGluPheLeuGln-----LeuLysLysValTh 1243

Db 5440 AAAGAACCAATTCCTACGAAGCGAATGAACGTAACACCAATCGCTCAACGCAAGAC 5499  
QY 1243 rLeuValValGlnLysArgArg 1251  
Db 5500 ATTACTGCTGAAGAATTCGAGCGGCTTAATGCGAAGTAGAATGCTGTGACACCAAGCA 5559  
QY 1252 ---AlaLeuLeuGlnMet 1262  
Db 5560 AATAGCAACACTGAAGCTGCTAATAGTCAAAATGATAGACCAACGCGAAGACACAG-- 5617  
QY 1262 uTyrLeuHisLeuArgGluValThrLys 1277  
Db 5618 ---GTGAAATAGTATTGATCAAGTACCAACACAGTAAATAAAGCAACT 5667  
QY 1277 eHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaLava 1297  
Db 5668 GCACGTAATGAATCAACAGCAATTT 5700  
QY 1297 lSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArgAsnSerPh 1317  
Db 5701 AAATTGCAAGAGATTCAAGCTACGCCAG-----ATGCAACAGATGAAGAAACCAAGCA 5754  
QY 1317 eLeuGlnLeuArgGlnAlaAlaAlaThrLeuGlnArgTyrArgAlaLysLysAlaMe 1337  
Db 5755 GCTGATGCTGAAGCAATA-----CTGAAAT 5781  
QY 1337 tLeuLysGlnLeuLysSerTyrAlaGln 1346  
Db 5782 GGTAAAGCAATCAAGCCATTTTCAGCAGCAACTACTAACCCCAAGTGTGATGAAGCTAAA 5841  
QY 1347 ---LeuLysGlnAlaAlaAlaThrLeuGlnThrArgTyrArgAlaLysLysAlaMe 1364  
Db 5842 GCAAATGCAGACAGCGATTATGCGGTAAACCAAAAGTTGTGAAGAAACAGCGGCT 5901  
QY 1364 tGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaAlaLeuLysValGlnArgAr 1384  
Db 5902 AAAGATGAATTCATCAATTAC---AAGCAACGCAAAACAATGTTATCAATAATGATCAG 5958  
QY 1384 gTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyr-----GlnLysGI 1401  
Db 5959 A-----ACGCTACAGAGAAAGAAAGACAGCTATTCAACAATTAGCAACAGCA 6009  
QY 1401 n-----ArgGlnAlaValIleArgLeuGlnLysTyrTyrArgSerIleArgaspMe 1418  
Db 6010 GTTACAGACCGGAAATAATATATACAGTGCAGTGCAGTGC-----AT 6048  
QY 1418 tArgLeu-CysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgL 1438  
Db 6049 GATAATGGTGTAGATCAGGC-----GAAAGAC 6075  
QY 1438 ystTyrArgAla-ThrValGln-AlaArgArgGlnArgGlu----- 1450  
Db 6076 GCTGGAAGAAATCAATTCACAGCGCAACCAACAGCGGTTAAATCAATGCTAAA 6135  
QY 1451 -IlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLe 1470  
Db 6136 AATGATGTTGATCAAGCTGCACTCAAAATCAAGCAATTTATATACAACTGGTGT 6195  
QY 1470 uLeuMetArg-----GlnGlnArgArgGluPheGluMetLysArgArgAlaLava 1487  
Db 6196 ACAACTGAAGAGAAAAAATGAGCAAAAGATTTAGTTTAAAGCTAAAGAAAAAGCGTAT 6255  
QY 1487 lValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGI 1507  
Db 6256 CAAGATATCTTAATGCACAAA---CAACTAATGATGTTACGCAAAATTAAGATCAAGCA 6312  
QY 1507 nLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMet-- 1526  
Db 6313 GTTGCTGATATTCAGGTATTTACTGCAGATACAACTAAAGATTTGCGAAGATGAA 6372  
QY 1527 -----LysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLe 1543

Db 6373 TTAGCAACAAAGCAAAACGAAACAAAAGCGCTTA---TTGCACAAATCGCAGATGCGACT 6429  
QY 1543 uGlnGlnLysPheArgGlyLysArgLeuMetIleGlnGlnArgAsnCysPheGlnLeuLe 1563  
Db 6430 ACTGAAGAAA---AAGAACAAGCAAAATCAACAAAGTAGACCAACAAATTAACACCAAGGTAAT 6486  
QY 1563 uArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPh 1583  
Db 6487 C-----AAAAT 6492  
QY 1583 eGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaLysVa 1603  
Db 6493 ATTGAAAATGACACGATCAATCGATGATTAACACTGCAAAAGATATGCAATTCACGCA 6552  
QY 1603 lIleGlnArgTyrTyrArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLe 1623  
Db 6553 ATTGACCCCAATTC-----AAGCATCAACAGATGTT 6582  
QY 1623 uAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValAr 1643  
Db 6583 A-----AAACGAATG-----CAAGAGCGGAATGC-----TAACGTGA 6615  
QY 1643 gCysLysValGlnGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLe 1663  
Db 6616 ATGCAAAATAAAATACTGAAATACTTAATAATAATGAGACTACTAATGAAGAAAAAGGT 6675  
QY 1663 uAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTyrCy 1683  
Db 6676 AAGCATTTGGACAGTTAGACGCGCATATGAAGAAGGTTTAAATTAATATTATG----- 6730  
QY 1683 sSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVa 1703  
Db 6730 ----- 6730  
QY 1703 lAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSe 1723  
Db 6731 ---CAGCAACTACTACAG---GTGATGTAACTACTGCTAAAGATACAGCAGTACAAAA 6783  
QY 1723 rThrThrValAsnThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeuAr 1743  
Db 6784 G-----TTCAACAACATTCATG-----CAAACTCTGTTAAG 6813  
QY 1743 gTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyrValPheAl 1763  
Db 6814 A-----AACCAGCAGCTAAAAAGAAATAGATCAAGCTGCAGCTGATAAGAAA 6861  
QY 1763 aHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIleTy 1783  
Db 6862 ACACAAATAGACAAACACCAAAATGCATCACAACAAAGAAA----- 6901  
QY 1783 rMetValArgGluThrLysLysLeuValAlaAlaArgLysGluLysMetLysGln 1800  
Db 6902 -TTAATGATGCAAAACAGAAAGTTGATGACTGAATTAATCAACGGAACAA 6952

## RESULT 15

US-09-976-594-907  
: Sequence 907, Application US/09976594  
: Patent No. 6673549  
: GENERAL INFORMATION:  
: APPLICANT: Buchbinder, Jenny  
: TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
: FILE REFERENCE: PA-0041 US  
: CURRENT APPLICATION NUMBER: US/09/976,594  
: CURRENT FILING DATE: 2001-10-12  
: PRIOR APPLICATION NUMBER: 60/240,409  
: PRIOR FILING DATE: 2000-10-12  
: NUMBER OF SEQ ID NOS: 1143  
: SOFTWARE: PERL Program  
: SEQ ID NO 907  
: LENGTH: 5923  
: TYPE: DNA  
: ORGANISM: Homo sapiens

FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: Incyte ID No. 6673549 977975.1  
 US-09-976-594-907

## Alignment Scores:

Pred. No.: 9,74e-14 Length: 5923  
 Score: 255.50 Matches: 291  
 Percent Similarity: 35.19% Conservative: 227  
 Best Local Similarity: 19.77% Mismatches: 531  
 Query Match: 2.69% Indels: 426  
 DB: 4 Gaps: 61

US-09-914-698-1 (1-1861) x US-09-976-594-907 (1-5923)

QY 252 ThrHisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThr 271  
 DB 941 ACACATGCGAGATGAAGTAATGTAAGTCACATCAGAGGGAGCATCA----- 988  
 QY 272 ProThrLysSerAlaIleHisAspLeuLys-ArgAspIleLysLeuValGlySerProLe 291  
 DB 989 ---AAATATTCCAGATTTCGAGATTGAAATGAAGATTGAAA-----CAGAGGTT 1036  
 QY 291 uArgLysTyrSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAl 311  
 DB 1037 GAGAAAAATTTCAGCAAGAACAAAGAAATACTT-----TTGGATAA 1075  
 QY 311 alileGlnGlySerMetProAsnLeuAenGlu-----MetLysIleArgSerIleGlu 329  
 DB 1076 AGTCATAGTTTACAGTTACAGCTGATGAGTGAAGATTATGTTGCTGATGATCTGAAAG 1135  
 QY 329 nAsnArgTyrTyrGlnGluGlnGlnIleLysAlaLysAlaLysAspLeuAsnSerSerSe 349  
 DB 1136 CGAGAGA-----GAAAGCTGAAGTCCCTTTTGGCAGCTAAAGAAAGCAACATGAAGA 1189  
 QY 349 rSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLe 369  
 DB 1190 AAGCTTAAGCACTATTGAGGCTCTGAAAAATAGATTAAATATTGAGAGTGCATCATTT 1249  
 QY 369 uAlaGlnSerSerArgPheAsn-----LeuHisGluValGlyArgLys 383  
 DB 1250 AGATCAGGAAGTCATTTCCAGTAACCGAAAGCAAGATATGCTTCTTAAACAAAGGTCAGAT 1309  
 QY 383 sSerValLysGlySerProValLysAsnPro-----HisLysArg----- 396  
 DB 1310 GTATATGGCAGACTCACAGTGTACTTCCCAGGTATACCAGGCCATATCCAAAGCAGATC 1369  
 QY 397 -----ArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrAr 414  
 DB 1370 TATGTTAAGACCTCTGGAACACTATCTTTA-----CCAGTCAAAACGTCATCTCTGA 1420  
 QY 414 gAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr 434  
 DB 1421 AAATGAATTTA-----AAGAAAGATTAGAACCAATGCGAATCGGAACCTTCTG 1465  
 QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTyrProHi 454  
 DB 1466 TGAGTCAGCAAAACAAGACCGACTGAAGCTCCAAAATGAACCTGGCACAC----- 1514  
 QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
 DB 1515 -----AAAGTGGCAGAAATGCAAAAGCTTTAGCATAGTAATGTGAAAG 1555  
 QY 474 rProArgLysValArgAspThrSerIleGlnPro-----Se 486  
 DB 1556 GGTCAAGAGAGATTGATGATGAACAGATGAAGCAATAGAAAGATGCATTAAGATGTGCA 1615  
 QY 486 rValLysLeuTyrAspSerGlu-----LeuTyrMetGlnThrCysIleAsnProAs 503  
 DB 1616 GAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1661  
 QY 503 pProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGl 523

DB 1662 -----CATTTTCTTCCCTTAAAGAACACATTAACAAG 1693  
 QY 523 uGlnAlaValAspArgHisGlnAlaAspPheLysLysTyrLeuAsnAlaLeuValSerIl 543  
 DB 1694 TGAAGCAGCCTCAGGGAAT-----CACAGACTAACCCGAGGA 1729  
 QY 543 eProAlaAspLeuAspAlaAspLeuAsnAsnLys-----lleAspValGl 558  
 DB 1730 ACTGAAGGATCAGTTGAAAGACTTGAAGTAAATATATGAAGGTGCTTCAACAGCAAGTGG 1789  
 QY 558 YLysLeuPheAsnGlnValArgAsnLysGluLeuValValAlaProThrLysGluGluGl 578  
 DB 1790 GAAATTAAGAAACCAATCAACAAATGAGATAGTAGAAGAGTTTAAAGGGATGA 1849  
 QY 578 nSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLe 598  
 DB 1850 AGGCAAG-----CTGATAGAAGAAATAAAGCATTCACAGAAGGAAT 1891  
 QY 598 uPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGl 618  
 DB 1892 TAGTATGTGTAATCGAGCGAGAGAGAAAGAAAGGTTCACAGAGATGGAAGGCCA 1951  
 QY 618 nAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIl 638  
 DB 1952 GGCAAAAGAAATGCTCAGCG----- 1970  
 QY 638 eLeuGluLeuLeuLeuCysPheAsnProLeuTyrLeuArgLeuGluValValPh 658  
 DB 1971 -----AAGTTGGCCTTTCCATTCCAGC 1993  
 QY 658 eGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLe 678  
 DB 1994 TGAAAAATTTGAAAAATCAAGAGCTCA----- 2021  
 QY 678 uAsnArgLeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuTh 698  
 DB 2022 -----TTATCAATGAAGTGAATCAG-----AAAGCAAAAAAATAGT 2059  
 QY 698 rGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuPr 718  
 DB 2060 AGAAATAGGAAGAGAACATGAAAAA-----TCACTTAGTGAATTAGACAGTTAAAGAG 2113  
 QY 718 oPheLeuSerGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheVally 738  
 DB 2114 AGAACTTGAGATGTTAAGGCCAG-----CTTCTCAGCAC-----GTCAA 2155  
 QY 738 sLysSerProHisLysGluThrLysAspIleLeuArgPheSerSerGluLeuLeuAl 758  
 DB 2156 ACCAGAGGAACATGAAGAGTAAAGCAGATTAGACAGAAATCAGGAACTTGGGA 2215  
 QY 758 aAsnIleGlyAspIleThrArgGluLeuArgIleGlyTyrValLeuGlnHisArgGl 778  
 DB 2216 GAAGATCACTGAGTTAATCACTTGAAAAATCAGACACTACAAAAAGGAATGAA-----AA 2269  
 QY 778 nThrPheLeuAspGlu-----PheAspTyrAlaPheAsnLeuAlaValAspLeuAr 796  
 DB 2270 AGTTATTGTAATAAAGTCTCTCAAGGAGCAGACATACTTAACTTGAATGAA 2329  
 QY 796 gAspGlyValArgLeuThrArgValValGluValIle-----LeuLeuAr 811  
 DB 2330 AAATCATTTATGTTCTTTTAAAGTAAGTGAAGACATGAAAAAGTCACATGATGCAATTAT 2389  
 QY 811 gAspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAs 831  
 DB 2390 TGATGATCTTAATAGAAAGCTT-----TTAGATGTAAACACAAAAATATACAGA 2437  
 QY 831 nValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAl 851  
 DB 2438 AAAGAGTTGGAATGGAGAAATGCTACTCGAAATATCAGACTTAAAGTAGGATGTAAG 2497  
 QY 851 aAlaGlnAspIleVal-----AspGlyHisArgGluLysThrLeuSerLeuLe 867  
 DB 2498 CCGCTAGAAACTGTGTTGTACCTCTCGAAACATGAAAAAGAGATAATAGCTCTGA 2557

QY	867	uTrpGlnLeuIleTyLysPheArgSerProLysPheHisAlaAlaThrValLeuGI	887
DB	2558	ATCAATATATGTT-----GAACTTAAGAAACAGCTGTCGAACCTTAA	2599
QY	887	nLys-----TrpTrpArgArgHisTrpLeuHisValValIleGlnArgArgIleArgHi	905
DB	2600	GAATAAATGGTGAAGACCGAGAGAAATACACGCTCTCACATCTGAAACACACTAACTT	2659
QY	905	sLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHi	925
DB	2660	GAAGAAGATCATG-----	2672
QY	925	sGlnMetArgLysTyTyValLysLeuPheLysThrGluArgThrGlnAlaAlaIle	943
DB	2673	-----AGTAATACGATGTGCGAGTTAAACCCATGAAGAGGTTAAATGACACTGAATGA	2728
QY	944	IleLeuGlnLysPheThrArgArgTyTrpLeuAlaGlnLysGlnLeuTyTrpSerTyHi	963
DB	2729	CAGTTTAGCCAAACTAACAGAGAAATATTAGATGTGAAGAAAAATTTGAAGATATAAA	2788
QY	963	SerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGI	983
DB	2789	T-----CA 2791	
QY	983	nArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgIle	1003
DB	2792	GGATTTTGTAAATAAAGATGAAGATGAATGAAATATTAAAAAGA-----	2834
QY	1003	uPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGI	1023
DB	2835	-----AACCTGGAAAAACACTCAG---AA 2854	
QY	1023	nGlnAlaAlaAlaSerTyTrpIleGlnMetGlnTrpArgThrTyTrpGlnLeuGlyArgIleGI	1043
DB	2855	CCAAATAAAGCTGAGTACATCAG-----CTGSC 2884	
QY	1043	nArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSe	1063
DB	2885	AGACACACGAG-----GC 2896	
QY	1063	rLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAs	1083
DB	2897	AAAGATGAGCTCGTTAAGTCAG----- 2918	
QY	1083	nIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLys-----CysAsnAlaAs	1100
DB	2919	-----AGCATGAGAAAGGTGCGAGATAGTAATGCTGA 2950	
QY	1100	pTyriLeuAlaLeuArgSerSerValLeuLysValGlnAlaTyArgLysAlaThrIleGI	1120
DB	2951	AATCTTGCC-----AATACAGAAAAAGC----- 2975	
QY	1120	nMetArgIleAspArgAsnHisTyTrpSerLeuArgLysAsnValIleCysLeuGlnGI	1140
DB	2976	-----CAAGAGAGATTGTGACACTGCGATGC 3001	
QY	1140	nArgLeuArgAlaIleMetLysMetArgGlu-GlnArgGluAsnTyTrpLeuArg-----L	1158
DB	3002	CGAAATTAAGCCAGAGAAAGGAGCTGCACACAAATACAAAGATGCAATTAAGCTAAATA	3061
QY	1158	eUArgAsnAlaSerIleLeuValGlnLysArgTyArgMetArgGlnGlnMetIleGlnA	1178
DB	3062	TGCCCCAATTGTCAGCTTTGAGGATGCGAGAGAAAAATTTAAAGCAACAGAGAAAGACT	3121
QY	1178	sPArgAsnAlaTyTrpLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgA	1198
DB	3122	AAAGACCGAGTTATCAGAGCGACACACAAAGATATAGTGTGAGAAAGAGACTCAAGAA	3181
QY	1198	IaThr---LeuGlnMetArgArgGluArgLysAsnTyTrpLeuHisLeuGlnThrThrL	1217
DB	3182	AAACAAGCAAGAGATGACAGTTAAAGAGAGGAGATTTTACC-CTTCAGAAAGAGATTGA	3240

Qy 1217 ysarglileGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheL 1233  
: : : : :  
Db 3241 GAGATAAGACAGTCTCATTGAGAAGTCTCATGAAATGGAAGA----- 3284

Qy 1237 euGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetA 1257  
: : : : :  
Db 3285 --GCATTAAACGCAAAAACAGACGACTAACAA-----CAGTTAA 3324

Qy 1257 rgLysGluArgGlnGlnTyrrLeuHisLeuArgGluValThrIleLysLeuGlnArgArp 1277  
: : : : :  
Db 3325 AAGACTTGTCACAGAAATACACCGAAGTAAGAATGTGAAGAGAAGCTAGTAGAAGAA- 3383

Qy 1277 heHISalagInLysSer-MetArgPheMetArgAlaLysTyrrArgGlyThrGlnAlaLa 1296  
: : : : :  
Db 3384 --AATGCCAAACAGACTTCTGAGATATCTTCGACTGCAG----- 3420

Qy 1297 ValSerCysLeu-----GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArg 1314  
: : : : :  
Db 3421 ATCTTTTTGCAAAACAAACATGTTCCATGGAAACAGGTGAGGCTCTG-AANAATCTCTT 3479

Qy 1315 AsnSerPheLeuGln--LeuArgGlnAlaLaIleThrLeuGlnArgArgTyrrArgAla 1333  
: : : : :  
Db 3480 AATGGCACAAATGAAATCTTAAGGAAGAACTGAAGAGTATGCAAAGTGTTCAGAGAAA 3539

Qy 1334 ArgLeuAsnMetIleLysGlnLeuLysSerTyrrAlaGln--LeuLysGlnAlaLaIle 1352  
: : : : :  
Db 3540 GAGCAGCAGACAGTGACCAACTGCATCAATTGTTGGAGAACTCAAAAGAACTCTCTGTA 3599

Qy 1353 ThrIleGlnThrArgTyrrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrrGln 1372  
: : : : :  
Db 3600 CCCTGGCAGAGCAATTTGCAGATTTAAAGAGCAATTTGAGAAAGAGTT----- 3647

Qy 1373 LysGlnArgGluAlaIleIleLysValGlnArgTyrrArgGlyAsnLeuGluMetArg 1392  
: : : : :  
Db 3648 -----GGAATCATAAAGCCAGCTTGAGAGAAAAGGAGAGAAAGCCAAAAAC 3695

Qy 1393 LysGlnIleGluVal-----TyrGlnLysGlnArgGlnAlaValIle 1406  
: : : : :  
Db 3696 AAAATGGAAGAAGTCTCCAAACTTCAGTCGGAGGTTCAGAAATCAATAAGCAAGCATTTAAA 3755

Qy 1407 ArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLysCysLysAlaGlyTyrrArg 1426  
: : : : :  
Db 3756 AAATTAGAGACT-----AGAGAGGTAGTTGACTTTGTCTAAATATAAGCAACAATAAGT 3809

Qy 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLys-TripArgAlaThrValGlnAlaAr 1446  
: : : : :  
Db 3810 GATTGGGAGACACAGATTTCTAGCTTAAATCAAAAATTTGGCCAATCTGTAATAGAAAGT-- 3867

Qy 1446 gArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIl 1466  
: : : : :  
Db 3867 ----- 3867

Qy 1466 eArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgArgAlaAl 1486  
: : : : :  
Db 3868 -----ATGAGGAAGATGTGAGGAAGTTTGGCATGCCAAAAGA----- 3906

Qy 1486 aValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTy 1506  
: : : : :  
Db 3907 -----AGAAATATCTGAAAAGATGAGAGGAATTA 3938

Qy 1506 rGlnLeuIleGlnSerServallIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526  
: : : : :  
Db 3939 CTGCAATTCAGCATTTGACCAAGAAATTAAG-GATCAGAAAGCAACGATGTGATAGTCCCT 3997

Qy 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLy 1546  
: : : : :  
Db 3998 A-----ACACAATCAAGATTTACAAAGAG 4024

Qy 1546 sPheArgGlyLysArgLeuMetIleGlnArg-----As 1558  
: : : : :  
Db 4025 AATACAGAAATCTGCTAAACAAATAGACCAAAAGATAATAGATACTGAAGTCTCTTAA 4084

Qy 1558 nCysPheGlnLeuLeuArgCysSerMetProGlyPheGln-----AlaArgAlaArgGl 1576

```

Db      4085  TGAATGGAAGATTAAACAGGCACTCAATGGCCCTTCCCAACTCACCCTACACAGTGG 4144
Qy      1576  yPheMetAlaArgLysArgPheGlnAlaLeuMetThr-----
Db      4145  GAAACCCACCAAGAGGAGCGAGCGCTGATTGACACTCTGCAGCACCAAGTGAATCTCT 4204
Qy      1589  -----ProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTy 1607
Db      4205  GGAGCAACAGCTGGCCGATGCTGACAGACAGCACCAGAAAGTAATTGCAATT----- 4256
Qy      1607  rTTPArgGlyTyrLeuIleArgArgArgGlnLysHis-----GlnGlyLeuLeuAspI 1625
Db      4257  -TATCGGACACACCTTCTTAGTGTGTCACAGGGTCAATGGATGAAGATGTTTCAGGAGGC 4315
Qy      1625  eArgLysArgIleAlaGlnLeuArgGln 1634
Db      4316  TCTGCTCCAGATCATACAAATGCGCAG 4343

```

Search completed: September 15, 2004, 13:05:22  
 Job time : 1990 secs



*This Page Blank (usp10)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 15, 2004, 06:57:11; Search time 1833 Seconds

(without alignments)  
5108.382 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVNSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODX1=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPIO\_spoolh/US09914698/runat\_10092004\_150008\_28369/app\_query.fasta\_1.2055  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09914698 @CGN\_1 1260 @runat 10092004\_150008\_28369  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match Length	ID	Description
------------	-------	--------------	----	-------------

1	834	8.8	2453	13	US-10-342-887-227	Sequence 227, App
2	834	8.8	2453	13	US-10-172-118-227	Sequence 227, App
3	707.5	7.4	3887	13	US-10-342-887-1746	Sequence 1746, App
4	707.5	7.4	3887	13	US-10-172-118-1746	Sequence 1746, App
5	700.5	7.4	4027	17	US-10-115-635-64	Sequence 64, Appl
6	678	7.1	2811	16	US-10-108-260A-956	Sequence 956, App
7	437	4.6	2414	13	US-10-342-887-1748	Sequence 1748, App
8	437	4.6	2414	13	US-10-172-118-1748	Sequence 1748, App
9	437	4.6	2414	17	US-10-188-832-81	Sequence 81, Appl
c 10	348.5	3.7	3069	17	US-10-437-963-53524	Sequence 53524, A
11	322	3.4	7677	10	US-09-814-353-21993	Sequence 21993, A
12	322	3.4	14800	9	US-09-954-456-1601	Sequence 1601, App
13	322	3.4	14800	15	US-10-269-909-61	Sequence 61, Appl
14	322	3.4	14800	17	US-10-717-597-183	Sequence 183, App
15	317.5	3.3	15231	9	US-09-917-800A-1505	Sequence 1505, App
16	317.5	3.3	15231	12	US-10-152-319A-1780	Sequence 1780, App
17	316.5	3.3	14835	15	US-10-240-965-113	Sequence 113, App
18	316	3.3	6401	15	US-10-171-581-352	Sequence 352, App
19	316	3.3	6401	15	US-10-177-293-314	Sequence 314, App
20	293	3.1	5229	16	US-10-012-697-1457	Sequence 1457, App
21	283.5	3.0	6946	10	US-09-814-353-20816	Sequence 20816, A
22	283	3.0	6839	13	US-10-342-887-1803	Sequence 1803, App
23	283	3.0	6839	13	US-10-172-118-1803	Sequence 1803, App
24	276.5	2.9	5457	13	US-10-282-122A-28061	Sequence 28061, A
25	275	2.9	5895	16	US-10-389-493-26966	Sequence 26966, A
26	273.5	2.9	5418	13	US-10-282-122A-27329	Sequence 27329, A
27	273.5	2.9	8503	15	US-10-037-270-130	Sequence 130, App
28	273.5	2.9	8503	16	US-10-117-722-130	Sequence 130, App
29	273.5	2.9	580073	15	US-10-205-220-1	Sequence 1, Appli
30	273	2.9	7434	9	US-09-815-242-4761	Sequence 4761, App
31	273	2.9	7437	9	US-09-815-242-8869	Sequence 8869, App
32	273	2.9	8493	14	US-10-071-766-51	Sequence 51, Appl
33	271.5	2.9	6075	16	US-10-388-934-532	Sequence 532, App
34	271.5	2.9	6075	17	US-10-664-705-62	Sequence 62, Appl
35	271.5	2.9	7992	10	US-09-893-519A-140	Sequence 140, App
36	271.5	2.9	8491	15	US-10-133-013-260	Sequence 260, App
37	270	2.8	11677	15	US-10-082-830-134	Sequence 134, App
38	268	2.8	10190	9	US-09-864-864-232	Sequence 232, App
39	268	2.8	10211	9	US-09-954-456-1153	Sequence 1153, App
40	268	2.8	10211	9	US-09-967-768A-186	Sequence 186, App
41	268	2.8	10211	10	US-09-918-624B-2	Sequence 2, Appli
42	265	2.8	10281	15	US-10-084-817-361	Sequence 361, App
43	265	2.8	10281	15	US-10-240-965-201	Sequence 201, App
44	260.5	2.7	11736	16	US-10-291-265-218	Sequence 218, App
45	260	2.7	6452	15	US-10-205-823-418	Sequence 418, App

## ALIGNMENTS

### RESULT 1

US-10-342-887-227  
; Sequence 227, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14

```

; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 227
; LENGTH: 2453
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-227

Alignment Scores:
Pred. No.: 3,21e-72 Length: 2453
Score: 834.00 Matches: 235
Percent Similarity: 44.15% Conservative: 153
Best Local Similarity: 27.04% Mismatches: 228
Query Match: 8.77% Indels: 254
DB: 13 Gaps: 22

US-09-914-698-1 (1-1861) x US-10-342-887-227 (1-2453)
QY 522 AspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuVal 541
DB 6 GATGAAGCTGGAAGGAAGGAAGCAGGCTTCACTTGGTGTAAATTTATATTA 65
QY 542 SerIleProAlaAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560
DB 66 ACC---CCTGATGACTTCACTGTAAACAAATATTTCTGAAGTAAATGCTGCTACTCTT 122
QY 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577
DB 123 CTTTGGGAATAGAGATCAACATCAATATAATAGTGTCTCTAGACGACCTACAAAGAGAA 182
QY 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLysArgLysAlaAlaVal 596
DB 183 ATGTCCTCAGAGCTTATACGCTCGGTGTAGGTAAACAGACTACGCTCGCAGCATGC 242
QY 597 GluLeuPheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616
DB 243 CGTTTGTGTACTTCTGAAAAATAGTTAAAGCTATTAAAGCTTGAATTTGAAATGAA 302
QY 617 LysGlnAlaLeuArgIleArgSerArgAsnLysHisLeuAspValValMetGlnArg 636
DB 303 GCTAGGGCGGTAAATGTTGGAAGATAGACACCTATGGAAGATGTGGGAGAACGTCAG 362
QY 637 ThrIleLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGluVal 656
DB 363 AAAGTCCTGAATTTGGCTGTGCTCAATCTTTGTGGCTTCAATTTGGTCTAGAGACA 422
QY 657 ValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676
DB 423 ACTTATGAGAACTCATATCTTTGGAAGATAACAGTGTACACAGGGTGGCTATGTT 482
QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGlnArgTyrSerLys 693
DB 483 ATTCTGAATCGCTACTTTTGAATCCTGATATAGCAGCTGATATAGACACCCACTGTT 542
QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle 713
DB 543 CCTCACCTGATAGAGATGTCATGAAGAGCTTTGTCCAAAGTTTACATTGAAAGAGTGA 602
QY 714 LeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733
DB 603 TTGTTGTGTGCTGTTTCTTGTGATTTATGCTTAAATTTCCAGACTCATTTGATGATCCT 662
QY 734 CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 753
DB 663 TGCTCTCTGTAAGATGCGGAATTCAGGCTAGTAAAGAAATCCCTTAGGCTTTTCA 722
QY 754 SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrVal 773
DB 723 CGAGATTTCCTAAAGTGGTGAAGGTGACCTTCCCGTCCGCTTGGCTTATTTGGGATTA 782
QY 774 LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal 793
DB 783 GTTAACCAATGTTACAGACCACTTTGATGAATTTGATTTTTCGCGTTACAAATCTTCGCGTA 842

```

```

794 AspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAsp 813
843 GACTTCCAATGTGGAGTGGCCCTTGTGCGAACCCATGGAACCTTCTCAGCAAGCTGGGAC 902
814 LeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLys 833
903 CTCTCAAGAAACTCAGGATTCGGCAATAAGTCGCTTCAAAAGATGCAATGTTGAC 962
834 LeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAsp 849
963 ATTGTTCTCAAGTTCCTTAATCACAGGAATTAAGTATGATGAGCATGGAATACA 1022
850 IleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGln 869
1023 ATTCTATCTAAGGATATTGTGGATAGGCAAGAGAAAAAACTCTCAGGTTGCTTTGAAA 1082
870 LeuIleTyrLysPheArg----- 875
1083 ATAGCGTTTGTCTTTTTCAGGTGGATATTTCCTTAACCTTAGATCAATTAAGGAAGAAATT 1142
875 ----- 875
1143 GCCTTTCTAAACACACAAAGAGTATAAGAAAAACAATATCTCTACTATCATGCCATTCT 1202
876 -----SerProLysPheHis----- 880
1203 GATGATCTTATTATAAGAAAAAGCAAAAGGATAGTGGTCTCTTGAACAATATAGT 1262
881 -----AlaAlaAla-ThrValLeuGlnLys 888
1263 GAAACATAAAGTTATTGATGGATGGGTAATATGCTGTTGTGGCTTCTTATAATAAAG 1322
888 sTrp----- 889
1323 GTGGAGATTTTACAGTGTCTTCTCAGACGGCGTGTGTATGTATTACTGATCCACCAT 1382
889 ----- 889
1383 TACCATCCTTGCTATGTGCCATTAGAGCTTATATGTCAGCGTACTACTCAAACTGTGGAA 1442
890 -----Tr 890
1443 TGACGCAAACTGGTTCAGTGGTATTAAATTCATCATCTGAACTGTATGATGAGTTCCTG 1502
890 pArgArgHisTrpLeuHisValVal----- 898
1503 GATATGTCATTAAGCAATTTGATCATGAAATATCTCAGAGCTATACAAAGAGCTCCTA 1562
899 -----IleGlnArgArgIleArgHis----- 905
1563 GAAATGAAAGAAAAAATTTT-CACCTTGTGTAGGTCTGCGATTAGAGACCTTGGTGGAA 1621
905 ----- 905
1622 ACCTGCTATGATTAATCAATTCAGANATGTCAATATACAAATCCAGATGAAAGGTGGTTAT 1681
906 -----LysGluLeuMetArgArgHisArgAlaAl 915
1682 TACCTATTGTCATTTCTTTGTGCAAGCTTTTGGATCTTCGTAAGAAATAAGAGCTGC 1741
915 aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLys 935
1742 TCGACTCATCAACACACATGGAGAAATATATAAATAAACAAGATCTCAACGCCATCA 1801
935 sThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaGln 955
1802 G---GAGAGAGAGAAAGCTGCAAGAAATATTCAATTCGCTGTTGATCAATTTCTAGCAA 1858
955 nLysGlnLeuTrpGlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGln 975
1859 ACAAGATTGAGAAAAAGAGTTAATGCGAGCACTCGTCAATTCAGAAATATTGG----- 1910
975 nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLe 995

```

Db 1910 ----- 1910

QY 995 uGlnArgIleTTPArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015

Db 1911 -----CGAAGAGTCTTAGCAGAGAGAAATATTAAATGTTAAAGAGGA 1954

QY 1015 aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035

Db 1955 AAAGCTGGAAAAGTTCAAAATAA-----CGAGATCACATTATCAGGATATGGAG 2008

QY 1035 gThrTyrGlnLeuGlyArgIleGlnArgHISgluPheLeuArgGlnArgAspLeuIleMe 1055

Db 2009 AAGATATTCC-----ACTAGACAAAGATTCTCGAAATGGAATATTATTCAAT 2056

QY 1055 tPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGluPheGl 1075

Db 2057 CATCTGCAATCTAGATAAGA-----ATGATAATTGCTGTATCATCTTATAA 2104

QY 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetAr 1095

Db 2105 ACGATATCTTTGGGTACAGTTACAAATTCAGAGGCAATGGCGTCTTATTAAAGAGAA 2164

QY 1095 GlyCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAla---Ty 1114

Db 2165 ACAAGATCAACAAGATATGAAATGCTAAATCATCACTCTTATAATCCAAATCTATGT 2224

QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134

Db 2225 CAGAAAA-----TGGAAGCAACGTAFAA-- 2246

QY 1134 nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGlnArgGl 1153

Db 2247 -----ATGCATCAAGTAAGTAAGCTACAGTANATTCGAAAGCTTTAGAGA 2296

QY 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169

Db 2297 ATGGCAATTAAGAAAAACAAGCTAAAGAGAAATTCCTGCTATATATCATCAATCATGTA 2356

QY 1169 rArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIl 1189

Db 2357 TAGAATGCATAAGAAATTA-----CGGAAGTATATATTTATATAGATCTTGTGT 2404

QY 1189 eIleAsnValGlnArgArgTrpArg 1197

Db 2405 TGTTATCATTCAGAAAGATTTCGG 2429

## RESULT 2

US-10-172-118-227

; Sequence 227, Application US/10172118

; Publication No. US20030224374A1

; GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue

; APPLICANT: He, Yudong

; APPLICANT: Linsley, Peter

; APPLICANT: Mao, Mao

; APPLICANT: Roberts, Chris

; APPLICANT: Van 't Veer, Laura

; APPLICANT: Van de Vijver, Marc

; APPLICANT: Bernards, Rene

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

; FILE REFERENCE: 9301-175-999

; CURRENT APPLICATION NUMBER: US/10/172,118

; CURRENT FILING DATE: 2002-05-14

; PRIOR APPLICATION NUMBER: 60/380, 770

; PRIOR FILING DATE: 2002-05-14

; NUMBER OF SEQ ID NOS: 2699

; SEQ ID NO 227

; LENGTH: 2453

; TYPE: DNA

; ORGANISM: Homo sapiens

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: AK001380

; DATABASE ENTRY DATE: 2001-06-18

US-10-172-118-227

## Alignment Scores:

Pred. No.: 3,21e-72 Length: 2453

Score: 834.00 Matches: 235

Percent Similarity: 44.65% Conservative: 153

Best Local Similarity: 27.04% Mismatches: 228

Query Match: 8.77% Indels: 254

DB: 13 Gaps: 22

US-09-914-698-1 (1-1861) x US-10-172-118-227 (1-2453)

QY 522 AspGluGlnAlaValAspArgHISgluAlaAspPheLysLysTrpLeuAsnAlaLeuVal 541

Db 6 GATGAACCTCGAAGCAAGCAAGCAAGCGCTTCACCTGGTGGTAAATTTATATTA 65

QY 542 SerIleProAlaAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560

Db 66 ACC---CCTGATGACTTCACCTGTAACAAATATTCTGAAGTAAATGCTCTACTCTT 122

QY 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577

Db 123 CTTTGGGAATAGAGAAATCAACATAAAATAAGTGTTCCTAGAGCACCTACAAAAGAGAA 182

QY 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaVal 596

Db 183 ATGCTCTCAGAGCTTATACCTGCTGGGTAGTTAAACAGACTACGTCGTGACAGATGC 242

QY 597 GluLeuPheSerGlnGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616

Db 243 CGTTTGTTCCTTCGAAAAAATGGTTAAAGCTATTAAAGCTTGAATTCGAATGAA 302

QY 617 LysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArg 636

Db 303 GCTAGGCGGTAAATGTTGGAAGATAGACACCTATCGAAAGATGTGGAGAACGTCAG 362

QY 637 ThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLysGluLeuVal 656

Db 363 AAAGTCTCGAANTGGCTGTTCCTCAATCCTTTGTGGCTTCGAATTTGGTCTAGAGACA 422

QY 657 ValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676

Db 423 ACTTATGGAGAACTCATATCTTTGGAAGATAAACAGTATGTCACAGGTTGGCTATGTT 482

QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693

Db 483 ATTCTGAATCGCTACTTTGGAATCCTGATATAGCAGCTGAGTATAGACACCCACTGTT 542

QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle 713

Db 543 CCTCACCTGTATAGAGATGGTCATGAAGAAGCTTTGTCCAAAGTTTACATTGAAAAGTTA 602

QY 714 LeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733

Db 603 TTGTTGTTGGTCTCTTTCTTGATATGCTAAATTTCCAGACTCATTCATCATGATCCT 662

QY 734 CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 753

Db 663 TGTCTCTCTGTAAAGATGCCAATTCAGGCTAGTAAAGAAATCCTTTAGGCTTTTCA 722

QY 754 SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrVal 773

Db 723 CGAGATTTCCTAAGTGGTGAAGGTGACCTTTCCCGCTCGCCTTATTTGGGATTAACCT 782

QY 774 LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal 793

Db 783 GTTAACCATGTTCCAGACACCATTTGATGATTTGCGGTTACAAATCTTCGCCGTA 842

QY 794 AspLeuArgAspGlyValArgLeuThrArgValGluValIleLeuLeuArgAsp 813

Db 843 GACTTGCAATGTGGAGTGGCGCTTGTGCGAACCATCGAACTTCTTCACACAGAACCTGGAC 902



## Alignment Scores:

Pred. No.: 2,89e-59 Length: 3887  
 Score: 707.50 Matches: 276  
 Percent Similarity: 40.70% Conservative: 225  
 Best Local Similarity: 22.42% Mismatches: 412  
 Query Match: 7.44% Indels: 319  
 DB: 13 Gaps: 40

US-09-914-698-1 (1-1861) x US-10-342-887-1746 (1-3887)

QY 868 TrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln 887  
 DB 359 TGGATGATGAGAAAGAGATCGGA-----GAGATGACAGGGCTGCTACTTTCATCCAG 412  
 QY 888 LysTrpTrpArgGHisTrpLeuHis-----Val 897  
 DB 413 TCTACTTTTCAGAAATGACAGATATACATATGATGATCAGGCTTTGAAACAGGCGCTCGTT 472  
 QY 898 ValIleGlnArg-----ArgIleArgHisLysGluLeuMet 909  
 DB 473 GTGATCCACAGCAATACCAAGCAATAGAGCTGCAAAACTGACAGGCGAGCATATATCTC 532  
 QY 910 ArgArgGHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys 929  
 DB 533 AGACAAAGACACATCTGCTGTGATCCTTCAGGCTCATTCAGGGGTATGAAAACTAGAAGA 592  
 QY 930 TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThr 949  
 DB 593 CATTTGAAGAGTATGCAATTC-----TCTGCAACCCCTTATTCAGAGTAGGTTT 640  
 QY 950 ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisIleIleThrIleGln 969  
 DB 641 AGATCATTTACTGGTCAGAGAGAGATTCATTTCCCTCAAAAAGCTACTATTTTGTTCAG 700  
 QY 970 ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg 989  
 DB 701 AGGAAATATCGAGGCACCACCATTTGTGTCACAAATATAAATTTGTACCAATTTCTGCACTAAGA 760  
 QY 990 GluAlaAlaIlePheLeuGlnArgIleTrpArgArgGluLeuPheAlaLysLysLeu--- 1008  
 DB 761 AAGCAGCCATTTACAAATACATGCTCTTACAGAGAGCTGATGGTAAAGAGAGTTACAA 820  
 QY 1009 -----LeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln--- 1023  
 DB 821 GAATGCAAGGGCTGCAGT-TCTCATTCAGGCTACTTTCAGGATGCACAGAAAAA 879  
 QY 1024 -----GlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThr 1036  
 DB 880 TATATTACATTTGACATTTGAAACATGCTTCAATTTCTAATTCAGCAACATTTATCGAACA 939  
 QY 1037 TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe 1056  
 DB 940 TATAGAGCTGCAAAATTCGAAAGAGAAATATATATCAGCAATGGCATTCGCTGTGTT 999  
 QY 1057 ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGlu 1073  
 DB 1000 ATTCAGGCTGCATATAAAGAGATGAAGACAGCAAACTTTTAAAGGAAACACAAAGCT 1059  
 QY 1074 -----PheGlnGlnLeu 1077  
 DB 1060 TCTATTGTAATACAGGCACCTACAGAAATGTATAGGCAGTATTTGTTCTACCAAAAGCTT 1119  
 QY 1078 LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys 1097  
 DB 1120 CAGTGGGCTACAAAATCATCAAGAAAAATATAGAGCAAAATAAAAAAGAAACAGAAA--- 1176  
 QY 1098 AsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----ValGlnAla 1113  
 DB 1177 -----GTATTTCACCAATCACTTAAGAAAGAGACTTGTGTTTCAGGCA 1221  
 QY 1114 ---TyrArgLysAlaThrIleGlnMetArgIleAsp 1124  
 DB 1222 GGTTCAGGACATGAACATAAAAAACAGATTTCAGNACAGCAGCAGGCTGCCATTTATT 1281

QY 1125 -----ArgAsnHisTyrTyrSerLeuArgLysAsn 1134  
 DB 1282 ATTCAAGACATTGTAAAGCCCTTTAAATAAAGAGCATTTATCTCCACATTAGACCAACA 1341  
 QY 1135 ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln----- 1151  
 DB 1342 GTAGTTTCTATTCAAAGAAGATACAGAAACTTAACCTGCAGTCGCTACCAAGACAGTTATT 1401  
 QY 1152 -----ArgGluAsnTyrLeuArgLysArgAsn 1160  
 DB 1402 TGTATACAGTCTTATTACAGAGGCTTTAAAGTACAGAAAGATATTCAAAAATATGCACCGG 1461  
 QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
 DB 1462 GCTGCCACATAATTTCAGTCAATTCATGCAATGACAGAGGCGGCAAAAGTTGAT--- 1512  
 QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTyrArgAlaThrLeu 1200  
 DB 1513 ---TATGAAACAAAGAAACTGCAATTTGTGTATACAGAAATTTATAGGTTGTATGTT 1569  
 QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln 1220  
 DB 1570 AGAGTAAACAAACAGAAAGAAACTTTTACGAGTTTCAAAATCTGTACGAACATTTCTCAG 1629  
 QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
 DB 1630 GCTGCTTTTAGAGGCGATGAAA-----GTTAGACAAATAATTGAAAAATGTATCAGAGAA 1683  
 QY 1241 LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 DB 1684 AAGATGCGACCATTTGTT-----AACCAATCTGCACCTGCTGTGTACAGAGTAAA 1734  
 QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
 DB 1735 ACTCAGTATGAAGCTGTTCAAAAGTGAAGGTGTATGATTCAAGAGTGTGTATATAAGCTTCT 1794  
 QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
 DB 1795 GCGCTTGCTGTGTTTCAGAGACAGAGATCATCTTCAAAGTAGGCTGTGAGTAAACAATT 1854  
 QY 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu----- 1313  
 DB 1855 CAAAAAGCTTTTGTAGAAATGCTCACAAGAAACTGGAACACACAGAAATGTGCTGCCCTA 1914  
 QY 1314 -----ArgAsnSerPheLeuGlnLeuArgGln 1322  
 DB 1915 CGGATTCAGTTCTTCTTCAGATGCTGTGTATTCGAGAGAGATTTGTTTCAGAGAAAGA 1974  
 QY 1323 AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLys 1342  
 DB 1975 GCTGCTATCCTTACAGCATTTATTTAGGAG-----TGCAAAACCAAGAAA 2022  
 QY 1343 SerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLys 1362  
 DB 2023 CAGTTTACTATATAGAAAGACAGCAGCTGGTGTTTTACAAAATCACTACAGAGCATTTCTG 2082  
 QY 1363 AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln 1382  
 DB 2083 TCTGCAAAACATCAAGACAGATCTTATTCAGATCAGACAGCGTGTATCATTTATCCA 2142  
 QY 1383 ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg 1402  
 DB 2143 GCTAGAGTAAAGGATTTATACAGAAACCGAAG-----TTTCAGGAATTTAA 2190  
 QY 1403 GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys 1421  
 DB 2191 ATATGACCATATAAATTCAGGCTATGTGGAGGAGATATAGAGCCAGAAATATTATTGT 2250  
 QY 1422 Lys-----AlaGlyTyrArg----- 1426  
 DB 2251 AAGTGAAGCTGCTGCAAGATTCAGGCTGTGTATAGATGTTGGAGAGCACACAAAGA 2310

QY 1426 ----- 1426  
 Db 2311 TATCTAGCTATATTAAGAGCTGTTAAATATTAATCAAGGTGCTTCTATACCAACTAGAG 2370  
 QY 1427 ----- 1427  
 Db 2371 AGAACCGCTTTTGAATGTGAGCATCATGCAATATATTCATTCAGAGAAATGGAGACT 2430  
 QY 1442 ThrValGlnAlaArgGlnArgGluLeuPheLeuSerThrIleArgLysValArgLeu 1461  
 Db 2431 ATACTTCCTGCAAGATAGCTCATGAACACTTCTTAATCATATAAAGA----- 2478  
 QY 1462 MetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMet 1481  
 Db 2478 ----- 2478  
 QY 1482 LysArgArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLys 1501  
 Db 2479 ---CATCGAGCTGCTGTTGATCAAGCACATATATAGAGGA-----TATAAA 2523  
 QY 1502 AlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArg 1521  
 Db 2524 GGAAGCAGCTCTCTTCGCGAGAAATCTGCTGTTGATCATACAAAAATATATACGA 2583  
 QY 1522 AlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaVal 1541  
 Db 2584 GCCAGGAGGCTGGAAGCATGAAGGATAAATATATGAATTTAAATAATCTACAGTT 2643  
 QY 1542 HisLeuGlnGlnLysPheArgGly-----LysArgLeuMetIleGlnGlnArgAsn 1558  
 Db 2644 ATCCTACAAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2703  
 QY 1559 CysPheGlnLeuLeuArg----- 1564  
 Db 2704 AAAATTCGACTTCTTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2763  
 QY 1564 ----- 1564  
 Db 2764 CAAAGAGCCTATAAACTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2823  
 QY 1565 ---CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
 Db 2824 ATCTGTATTCAGATGCTTTTGAGCAGAA-----TTCAAGAAAGAGATTT 2871  
 QY 1584 ---GlnAlaLeuMetThrProGluMetMetAsp-----LeuLeuArg 1596  
 Db 2872 ATTCAGAAATATCATAGCATCAAAAGATTCAGCATGAAGTCAAGAAATGCTGAGCCAG 2931  
 QY 1597 GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuLeuArgArg 1616  
 Db 2932 CGAAATAGGCTGCATCATGATATACAGAAAGCAGTGGCGCCATTTCTCTCGGTAAAG 2991  
 QY 1617 GlnLysHis-----GlnGlyLeuLeuAspIleArg----- 1626  
 Db 2992 CAGGAAATATTCATAGTGAATCATATAATTCAGGCATTATGAGAGGCTATTCTTGG 3051  
 QY 1627 -----LysArgIleAlaGlnLeuArgGlnGlnAlaLysAlaValAsn 1640  
 Db 3052 AGGAAGAAATGATTTGATCAAAATATAAGCTATATACGCTAAGTCTTCAAGTCTTAA 3111  
 QY 1641 SerValArgCysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheIleAlaSer 1660  
 Db 3112 AGG-----GAGATTCGAGAGAAACAACTCTACAAAGACTGCAGTTCGACTT 3162  
 QY 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu 1680  
 Db 3163 CAATACCTTTTGACATATAGACACCTTTCTGCGCAATTTCTGAGCCCTTAAACACCTAG 3222  
 QY 1681 MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg 1700  
 Db 3223 GTAGTTACTAGA-----TTGCTCCACTTGTGTGAGAACTGGCCAGAGTGGAGCA 3276  
 QY 1701 SerGluValAspLysGlnLeuLeuGluArgCysSerArg----- 1713

Db 3277 ATTTCTAAATATTTGTTTATCCGAGAGTTGTAATCGAGTATCTCTGTATGGAAGTC 3336  
 QY 1714 ----- 1714  
 Db 3337 ATCAGATATGCTGCGAAGCTTGTCTTATGATATCTAAGTATGAGAAACTACTTTCAGCA 3396  
 QY 1728 ThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
 Db 3397 GTTTATGATGAGAAAATTTATAGATATATATGAGCTTTTGCAGATATACCGAGAA 3456  
 QY 1743 -----ArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeu 1758  
 Db 3457 AAGCTGCTAATAAGTTGCGACAAAGCGGAGAGCATTTTACAAAACTTGTGTTG 3516  
 QY 1759 IleTrpValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsn 1778  
 Db 3517 TTGCTATTTTACTG-----AAGACAAACAAATAGAGCCCTCTGATGACGAGTAGG 3567  
 QY 1779 ProGluAlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMet 1798  
 Db 3568 TCCAAAGTTGTCACCGTATTTACAGTCTCTACAACTTACAGCTCATTAACATAAATG 3627  
 QY 1799 LysGlnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys--- 1817  
 Db 3628 ---AATACTGAAAGATA-----CTTTACAAAGCAAAAGAGAAAT 3663  
 QY 1818 -----IleAsnPheThrProCysSer-----LeuProSer 1827  
 Db 3664 TCTTCTAAGACATCTCTTTTATCCCAAGAACACCTGTGAAGCAGAGATAGTTTCAAGA 3723  
 QY 1828 LeuGluProAspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyr 1847  
 Db 3724 CTTAAGCCAGATTGGGTTTGTGAGAGAGATAACATGAAGAAATCACAATCCCTGCAA 3783  
 QY 1848 AlaPheAspThrIleLeuCysLysLeuGlnIle 1858  
 Db 3784 GCTATTCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3816

RESULT 4  
 US-10-172-118-1746  
 ; Sequence 1746, Application US/10172118  
 ; Publication No. US20030224374A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter  
 ; APPLICANT: Mao, Mao  
 ; APPLICANT: Roberts, Chris  
 ; APPLICANT: Van 't Veer, Laura  
 ; APPLICANT: Van de Vijver, Marc  
 ; APPLICANT: Bernards, Rene  
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
 ; FILE REFERENCE: 9301-175-999  
 ; CURRENT APPLICATION NUMBER: US/10/172,118  
 ; CURRENT FILING DATE: 2002-06-14  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 2699  
 ; SEQ ID NO 1746  
 ; LENGTH: 3887  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: NM\_018123  
 ; DATABASE ENTRY DATE: 2001-06-18  
 ; US-10-172-118-1746

Alignment Scores:  
 Pred. No.: 2,89e-59 Length: 3887  
 Score: 707.50 Matches: 276  
 Percent Similarity: 40.70% Conservative: 225  
 Best Local Similarity: 22.42% Mismatches: 412



Qy	1135	ValleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	115
Dy	1342	GTAGTTTCTATTCAAGAAGATACAGAAACTAACTCGACTGCGTACCCAGCAGTTATT	1401
Qy	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
Dy	1402	TGTATACAGTCTTATTACAGAGCTTTAAAGTACGAAGAGATATTCAAATATATGACCGG	1461
Qy	1161	AlaSerileLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
Dy	1462	GCTGCCACACTAATTCAGTCTATCTCATGCAATGCACAGGCCCAAGTTGAT-----	1512
Qy	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
Dy	1513	---TATGAAACAAAGAAACTGCAATTTGTTGTTATACAGAAATTATTATAGTTGTTATGTT	1569
Qy	1201	GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln	1220
Dy	1570	AGAGTAAAAACAGAAAGAAAAACTTTTAGCAGTTCAGAAATCTGPACCAACTATTTCAG	1629
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys	1240
Dy	1630	GCTGCTTTTAGAGGCATGAAA-----GTTAGACAAAATTTGAAAATGTTATCAGAGAA	1683
Qy	1241	LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
Dy	1684	AAGATGCGACCATTTGTT-----AACCAATCTGCACTTCTGCTGTTACAGAAGTAAA	1734
Qy	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgGppheHisAlaGln	1280
Dy	1735	ACTCAGTATGAAGCTGTTCAAGGTGAAGGTGTTATGATTCACAGAGTGGTATAAGCTTCT	1794
Qy	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeu	1300
Dy	1795	GGCCTTGCTGTTCCACGAGACGAGTATCATCTCAAGTAGGGCTGCAGTAAACAATT	1854
Qy	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
Dy	1855	CAAAAAGCTTTTGTAGATGGTTCACAGAAACTCGAAACACAGAAAAATGCTGCTGCCCTA	1914
Qy	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
Dy	1915	CGGATTCAGTCTCTCCCTCAGATGGCTGTGTATCGGAGAAATTTGTTACAGAGAAAAGA	1974
Qy	1323	AlaAlaIleThrLeuGlnArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLys	1342
Dy	1975	GCTGCTATCACTTTACAGCATTTATTTAGACG-----TGGCAAAACGAGNAAA	2022
Qy	1343	SerTyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
Dy	2023	CAGTTTCTACTATAGAAAAAGCAGCAGTGGTGTTCACAAAATCCTACAGCATTTCTG	2082
Qy	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
Dy	2083	TCGTGCAAAACATCAAGACAAAGTCTATTTACAGATCAGAGCAGTGTATCAATTATCAA	2142
Qy	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
Dy	2143	GCTAGAGTAAAGATTTATACAGAAACGGAAG-----TTTCAGGAAATTTAAA	2190
Qy	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
Dy	2191	AATAGCACCATAAAAATTCAGGCTATGTGGAGAGATATAGAGCCCAAGAAATATTATGT	2250
Qy	1422	Lys-----AlaGlyTyrArg-----	1426
Dy	2251	AAAGTGAAGCTGCCTGCAAGATTCAGCGCTGTATAGATGTTGGAGAGCACACAAGAA	2310
Qy	1426	-----	1426
Dy	2311	TATCTAGCTATATAAAGCTGTTAAATATTTCAGGTGCTTCTATATACCAACTAGAG	2370
Qy	1427	-----ArgIleArgLeuSerSerIleGlnArgLysTrpArgAla	1441



Best Local Similarity:	22.34%	Mismatches:	412
Query Match:	7.36%	Indels:	319
DB:	17	Gaps:	40
US-09-914-698-1 (1-1861) x US-10-115-635-64 (1-4027)			
QY	868	TyrGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln	887
DB	359	TGGATGATGAAGAAAGGATGCGA-----GAGATGCACAGGGCTGCTACTTTCATCCAG	412
QY	888	LysTrpTyrArgHisTrpLeuHis-----Val	897
DB	413	TCTACTTTCAGATGCACAGATTACATATACAGATATCAGGCTTTGAAACAGGCGCTCCGTT	472
QY	898	ValIleGlnArg-----ArgIleArgHisLysGluLeuMet	909
DB	473	GTGATCCACACAGCAATACCAAGCAAAATAGAGCTGCAAAACTGCAGAGCGCATTTATCTC	532
QY	910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929
DB	533	AGCAAGAAGACACTCTGCTGTATCTTCAGCTTCAGGTCGATTGAGGGTATGAAACTAGAGA	592
QY	930	TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThr	949
DB	593	CATTTCAGAGATGATGATCC-----TCTGCAACCTTATTTCAGAGTAGGTTT	640
QY	950	ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969
DB	641	AGATCATTTACTGGTGAGGAGAGATTCAATTCCTCCAAAGAGCTACTATTATTTGTTTCAG	700
QY	970	ArgTyrTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg	989
DB	701	AGCAATATCGAGCCACCACTTGTGCAACATATAAATTGTACCAATCTTGCTCTTAAGA	760
QY	990	GluAlaAlaIlePheLeuGlnArgIleTyrArgArgArgLeuPheAlaLysLysLeu---	1008
DB	761	AAGCGACCATTAACAATCAGTCATCTTACAGAGACTGTGCTGTAAGAAGAGATTACAA	820
QY	1009	-----LeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln---	1023
DB	821	GAAATGCCAAGGCTGCACT-TCTCATTCAGGCTACTTTCCAGATGCACAGAAAAA	879
QY	1024	-----GlnAlaAlaAspTyrIleGlnMetGlnTrpArgThr	1036
DB	880	TATATTACATTCAGACTTGGAAACATGCTTCAATTCATTCAGCAACATATACGACA	939
QY	1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe	1056
DB	940	TATAGAGCTGCAAAATGCAAGAGAAAAATTATATACAGCAATGCGCATTCGCTGTGTT	999
QY	1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGlu	1073
DB	1000	ATTGAGGCTGCATATAAAGGAATGAAAGCAAGCAACTTTTAAGGCAAAACCAAGCT	1059
QY	1074	-----PheGlnGlnLeu	1077
DB	1060	TCTATCGTAAATCAAGACCTACAGAATGTATAGGCACTATTGTTCTACCAAGCTT	1119
QY	1078	LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097
DB	1120	CAGTGGCTTACAAAATCATACAGAAAAATATAGAGCAAAATAAAGAAACAGAAA---	1176
QY	1098	AsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----ValGlnAla	1113
DB	1177	-----GTATTTCAACAATGAACTTAAGAAAGAGAGACTTTGTGTTTCAGGCA	1221
QY	1114	---TyrArgLysAlaThrIleGlnMetArgIleAsp-----	1124
DB	1222	GGTITTCAGACATGAACATATAAAGAAACAGATTTCAGGAACAGCAGCGTCCCATTA	1281
QY	1125	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1134
DB	1282	ATTGAGAAGCATTTGTAAGCCTTTAAATTAAGGAAGCATTAATCTCCACCTTAGAGCA	1341

QY	1135	ValIleCysLeuGlnArgGlnLeuArgAlaIleMetLysMetArgGluGln-----	1151
DB	1342	GTAGTTTCTTATTCAAAGAAAGATACAGAAAACTACGTCGATGCGTACCCAGCAGTTATT	1401
QY	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
DB	1402	TGTATACAGCTCTTATTACAGAGGCTTTAAAGTACGAAAGGATATTCAAAATATGCACCG	1461
QY	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
DB	1462	GCTGCCACACTAATTCAGTCAATTCATGCAATGCACAGCGCCCAAGCTTGTAT	1512
QY	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
DB	1513	---TATGAAACAAAGAAAACTGCAATTTGGTTATACAGAAATATTATAGTTGTATGTT	1569
QY	1201	GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln	1220
DB	1570	AGAGTAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	1629
QY	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLys	1240
DB	1630	GCTGCTTTTAGAGGCAATGAAA-----GTTAGACAAAAAATGAAAAATGTATCAGAGAA	1683
QY	1241	LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
DB	1684	AAAGTGGCAGCCATTTGTT-----AACCAATCTGCACTCTGCTGTGTACAGAAACAA	1734
QY	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln	1280
DB	1735	ACTCAGTATGAAGCTGTTCAAAAGTGAAGGTGTATGATTCAGAGTGGTATAAGCTTCT	1794
QY	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu	1300
DB	1795	GGCTCTGCTGTTTCACAGGAAGCAGAGTATCATTTCTCAAAGTAGGCTGCTGTAACAAT	1854
QY	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
DB	1855	CAAAAGCTTTTGTAGAATGTCACAGAAAACTGGAACACACAGAAATGTGTGCTGCCCTA	1914
QY	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
DB	1915	CGGATTCAGTTCTCTCCTTCAGATGGCTGTGTATCGGAGAGATTTGTTTCAGCAGAAAGA	1974
QY	1323	AlaAlaIleThrLeuGlnArgTyrArgAlaArgLeuAsnMetIleLysGlnLysLys	1342
DB	1975	GCTGCTATCACTTTACAGCATTTATTTTAGGAGC-----TGGCAAAACAGAAAA	2022
QY	1343	SerTyrAlaGlnLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
DB	2023	CAGTTTCTTACTATAGAAAAAGCAGCGTGTGTTTACAAAAATCACTACAGAGATTTCTG	2082
QY	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
DB	2083	TCTGCAAAACATCAAGACAGTCATTTACAGATCAGAGCAGTGTATTCATTATTCAA	2142
QY	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
DB	2143	GCTAGAGTAAAGGATTTATACAGAAACCGAAG-----TTTCAGGAAATATAA	2190
QY	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
DB	2191	ATAGCACCATAAAAATTCAGGCTATGTGGAGAGATATAGAGCCCAAGAAATATTATGT	2250
QY	1422	Lys-----AlaGlyTyrArg-----	1426
DB	2251	AAAGTGAAGCTGCCTGCAAGATTCAAGCTCGTATAGATGTTGGAGAGCACACAAGAA	2310
QY	1426	-----	1426
DB	2311	TATCTAGCTGTATTAAGCTGTATAAATATTATTCAGGTTGCTTCTATACCAAACTAGAG	2370



Db 131 AAACAATATTGCAATTAAGAGCAGCTGCACTATTATTATCCAGCAATGT----- 178  
 Qy 745 ThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThr 764  
 Db 179 -----TACCGTTCCAAAATAATAGCT----- 199  
 Qy 765 ArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPhe 784  
 Db 200 -----GCACAAAGAGAGAGATATATG----- 223  
 Qy 785 AspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGly---ValArgLeuThrArg 803  
 Db 224 -----CAGATCGCGCAATCTTGATCAAACTCAAGCA 256  
 Qy 804 ValValGluValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProIle 823  
 Db 257 TTTGTTAGAGGATACCTTCGGAAGCAGATG-----AGGTTACAAAGAAAGCTGTT 310  
 Qy 824 SerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsn 843  
 Db 311 ATTTCACTACGTCTATTTCAGATGAGAAGGCT----- 346  
 Qy 844 PheGlnLeuGlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThr 863  
 Db 347 -----CGGCAATATTAT 358  
 Qy 864 LeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAla 883  
 Db 359 CTG-----AAATGTATAAGCAATT 379  
 Qy 884 ThrValLeuGlnLysTrpTrpArgHisTrpLeuHisValIleGlnArgGile 903  
 Db 380 ATTGTCATTCAATATTACTAT-----CATGCATACAAAGCACAGGTCAATCAGAGG--- 430  
 Qy 904 ArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArg 923  
 Db 431 -----AGAACTTCTTGCAAGTGTACAGGCGGTACAGACTCTTCATCATACAAAGCA 484  
 Qy 924 GlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaIle 943  
 Db 485 GGTATAAGTACGCGCACTAATCAAA-----CAACAATCTATAGCTGCTCTT 532  
 Qy 944 IleLeuLysLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHis 963  
 Db 533 AAAATTCAGTCTGCTTTAGAGGCTATAATAAAGCGTAAATATCAATCTGCTTCAA 592  
 Qy 964 SerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGln 983  
 Db 593 TCTATAATAAGATTACAGAGATGTTACAGGCGGTACAGACTCTTCATCATACAAAGCA 652  
 Qy 984 ArgPheValGluLeuArgGluAlaIlePheLeuGlnArgIleTrpArgArgLeu 1003  
 Db 653 CATTTTGAAGCAAGACAGCTGTGATTTCCCTCCAGTCTGCTTATCGGCTGGAAG 712  
 Qy 1004 PheAlaLysLysLeuAlaAlaAlaGluThrAla-----Arg 1016  
 Db 713 GTTCGGAACACATTAGAGGGAACATCAAGTGCCTTCGAAGATTCACTGCTCTTTAGA 772  
 Qy 1017 LeuGlnArgSerGlnLysGln-----GlnAlaAlaSerTyrIleGlnMet 1032  
 Db 773 ATGGCCAGGCCAGAAACAGTTTAGATTGTTTAAACAGCAGCAGATTAGTCAATCCAGCAA 832  
 Qy 1033 GlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAsp 1052  
 Db 833 AATTTCAGAGCATGGACTCGCAGGAAGGAAGCAATGTATGGAGTATATTGAATCCGTCAT 892  
 Qy 1053 LeuIleMetPheValGln-----ArgArg----- 1060  
 Db 893 GCGGTACTGCTTCAATCTATGCGAAGGGAACAACTGAGAGACAGCTTCAAGG 952  
 Qy 1061 -----MetArgSerLysTrpSerMetLeuGlnArgLysGlu 1073  
 Db 953 CAACATAAATGTGCTATCATCATACAGTCTATATAGATGCAATGCAATGCAACAAAGAAAG 1012

Qy 1074 PheGlnGlnLeuLysArgAlaIleAsnIleGlnArgTrpArgAlaLysLeuSer 1093  
 Db 1013 TGGAAAAATCAATAAAAGCTGCTCTCTGATTCAAGATATTATAGGCTTACAGTATT 1072  
 Qy 1094 MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGln--- 1112  
 Db 1073 GGAAGAGACAGACATCAATTATTTGAAACAAAGCAGCTGTAGTAACCTTTACAGTCA 1132  
 Qy 1113 AlaTyrArgLysAlaThrIleGlnMetArgIle----- 1123  
 Db 1133 GCTTATCGTGTATGAAAGTGAGAAAAAGATTAAGGATTGCAACAAAGCAGCTCACT 1192  
 Qy 1124 -----AspArgAsnHisTyrTyrSerLeuArgLysAsn 1134  
 Db 1193 ATACAGCTCAAAATACAGAGCTTACAAACCAAAAGAAATATGCAACCTATAGAGTTCA 1252  
 Qy 1135 ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsn 1154  
 Db 1253 GCTATTATAATTACAGATGTTATCGAGGTATTAATAATTACAAACCATCAGCATAAGGAG 1312  
 Qy 1155 TyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyr-----ArgMet 1171  
 Db 1313 TATCTTAATTTGAGAAGACAGCAATTAATAATCCAAATCTGTTTATAGAGGTATTAGATT 1372  
 Qy 1172 ArgGln-----GlnMet 1175  
 Db 1373 AGAAGACATATTCAACATGACAGGCGGCGCCATTTTATTAAAGCCATGTTAAAAATG 1432  
 Qy 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArg 1195  
 Db 1433 CATCAGTCAAGATAAGTTTACCATACATACAGAAAGCAGCTATTGTTTATCAGTAAGA 1492  
 Qy 1196 TrpArgAlaThrLeuGlnMetArgGluArgLysAsnTyrLysLeuHisLeuGlnThrThr 1215  
 Db 1493 TGTAGACATATTATCAAGGTAAATGCGCGGTGAAAGTACCTGCAATTTTGAAGCT 1552  
 Qy 1216 ThrLysArgIleGlnIleLysPhe-----ArgAlaLysArgGluMetLysLys--- 1231  
 Db 1553 GTTAAGTCTTTCAGGCAAGTTTTAGAGGAGTAAAGATTAGACGGACTTTAGAAAGATG 1612  
 Qy 1232 -----GlnArgAlaGluPhe 1236  
 Db 1613 CAGACTGCAGCAACACTCATTTCAGTCAAACTACAGAGATACAGACAGCAACATACITTT 1672  
 Qy 1237 LeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMet 1256  
 Db 1673 AATAAGTTAAAGAAAAATAACAAAAACAGTACAGCAAAAGATACCTGGGCAATGAAAGAA 1732  
 Qy 1257 ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArg 1276  
 Db 1733 AACATACAAATTCGAAAGGTATAACAAACTGAGGCAATTCGTAAATATACATTCAAGCTATT 1792  
 Qy 1277 PheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAla 1296  
 Db 1793 TTTAGGGAAGAAAGCTAGAGACATTTAAAAATGATGCAT-----ATAGCC 1840  
 Qy 1297 ValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSer 1316  
 Db 1841 GCAACTCTCATTCAGAGGAGATTAGAACTCTAATGATGAGAAGAAGA----- 1888  
 Qy 1317 PheLeuGlnLeuArgGlnAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsn 1336  
 Db 1889 TTCTCTCTCTCAGAAAACTGCTATTGTTGATTCGAGAAAAATATCGGGCACTCTT--- 1945  
 Qy 1337 MetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThr 1356  
 Db 1946 TGTAACAAGCATCACTTACAGTTCCTTCAGGTACAAAATGCAAGTTTAAAAATCCAGTCA 2005  
 Qy 1357 ArgTyrArg-----AlaLysLysAlaMetGlnLysGln----- 1367  
 Db 2006 TCATACAGAAAGATGATGATAAGGAAGGATGCGAGAGATGCACAGGGCTGCTACTTTC 2065



Db 527 ATTATTCAAGCTAGAGTAAGATTATACAGAAACGG----- 565  
 Qy 1130 SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg 1149  
 Db 565 ----- 565  
 Qy 1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln----- 1166  
 Db 566 -----AAGTTTCAGAGAAATTAATAATAGCACCATAAAATTCAGGCTATGG 613  
 Qy 1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186  
 Db 614 AGGAGATATAGAGCCAGAAA-----TATTATGTAAGTG 649  
 Qy 1187 LysCysIleIleAsnValGln-----ArgArgTyrArgAlaThrLeuGlnMetArg 1203  
 Db 650 AAAGCTGCTGCAGATTCAAGCTGGATAGATGTTGGAGAGCA----- 694  
 Qy 1204 ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPhe 1223  
 Db 695 -----CACAAAGATATCTAGCTGTATTAAGCTGTAAATATTATCAAGTTGCTTC 748  
 Qy 1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLysValThr 1243  
 Db 749 TATACCAAACTAGAG-----AGAACACCGTTTTTGAATGTGAGAGCATCAGCA 796  
 Qy 1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263  
 Db 797 ATTATCATTCAGAGAAATGAGAGCTATCTCTGCAAGATAGCTCATCAACACTTC 856  
 Qy 1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMet 1283  
 Db 857 TTAATGATAAAAGA----- 871  
 Qy 1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeuGlnMetHis 1303  
 Db 872 -----CATCGAGCTGCTTGTGATCCCAAGCACAT 901  
 Qy 1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323  
 Db 902 TATAGAGCA----- 910  
 Qy 1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343  
 Db 911 -----TATAAGAGGCGAGGCTCTCTCTCGGAG----- 940  
 Qy 1344 TyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363  
 Db 941 -----AAATCTGCTGCTTGTATCATACAAATAATATATACGAGCCAGGAGGCT 988  
 Qy 1364 MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg 1383  
 Db 989 GGAAGCATGAAGAGATAAATATATGAATTTAAATAATCTACAGTTATCTCTCAACAGCA 1048  
 Qy 1384 ArgTyrArgLysAsnLeuMetArgLysGlnIleGluValTyrGlnLysGlnArgGln 1403  
 Db 1049 CTGTGCTGGTGGCTAGTACGAAAGATTTTGA-----CAGAGA 1093  
 Qy 1404 AlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLysLysAla 1423  
 Db 1094 GCCAAAATCGACTT-----CTTCACCTCACTCAGCT 1126  
 Qy 1424 GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal 1443  
 Db 1127 GCATATTATCACTGAAT----- 1144  
 Qy 1444 GlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
 Db 1144 ----- 1144  
 Qy 1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArg 1483  
 Db 1144 ----- 1144

Qy 1484 ArgAlaIleValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
 Db 1145 -----GCTGTTAGAAATCAAGAGCCTATAAACTTTTACCTGGCTGTGAAGATGCTAAC 1198  
 Qy 1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
 Db 1199 AAGCAGGTT-----AATTCAGTCATCTGTATTCAGAGATGGTTTCGAGCAAGA 1246  
 Qy 1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
 Db 1247 TTACAAAGAAAAGAGATTTATTTCAGAAATATCATAGCATC----- 1285  
 Qy 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
 Db 1286 -----AAAAGATTGAGCATGAAGTCAAGATGTTCTCGTAAAGCAGCA 1327  
 Qy 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
 Db 1328 AATAGGGCTGCATCAGTAATACAGAAAGCAGTGGCCATTTTCTCTCCGTAAAGCAG 1387  
 Qy 1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaLysVal 1603  
 Db 1388 GAAAAATTCACCTAGT-----GGAATCATTAAG----- 1414  
 Qy 1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeu 1623  
 Db 1415 ATTCAGGCATATGAGAGAGCTTATTTGGAGAGAAAGAAATGATGTACAAAATATAA 1474  
 Qy 1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg 1643  
 Db 1475 GCTATACGACTAAGTCTTCAAGTCTTAAATAGGAGATTCGAGAGAAAACAACTCTAC 1534  
 Qy 1644 CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663  
 Db 1535 AAAAGAACTGCACCTTGCACTTCACTTACCTTTGACATATAAGCACCTTTCTGCCATCTT 1594  
 Qy 1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1683  
 Db 1595 GAGGCTTTAAACACCTAGAGTAGTGTACTAGA----- 1627  
 Qy 1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703  
 Db 1628 -----TTGTCTCCACTTGTGTGAGAACATGGCCGAGAGTGAGCAATTTCTAAA 1678  
 Qy 1704 AspLysGlnLeuIleGluArgCysSerArg----- 1713  
 Db 1679 ATATTGTTTGTATCCGAAAGTTGTAATCGCAGTATTCTTGTATGGAAGTCATCAGATAT 1738  
 Qy 1714 -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
 Db 1739 GCTGTGCAAGTCTTGTCTTAATGTATCTAAGATGAGAAAACACTTACTCAGAGTTTATGAT 1798  
 Qy 1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
 Db 1799 GTAGAAAATGTAATAGATATATCTATTGGAGCTTTTGAGATATACCGAGAAAAGCTGT 1858  
 Qy 1743 ---ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVal 1761  
 Db 1859 AATAAAGTTGCAGACAAAGCGGAGCATTTTACAAAAACTTGTGTTGTTGTTGCTATT 1918  
 Qy 1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781  
 Db 1919 TTACTG-----AAGACAACAATAAGACCTCTGTATGACGAGTAGGTGCCAAGTT 1969  
 Qy 1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801  
 Db 1970 GTTGACCGTATTACAGTCTCTCAAACTTACAGCTCATAAACATAAATG-----AAT 2023  
 Qy 1802 AlaArgLysProProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817  
 Db 2024 ACTGAAGATA-----CTTTACAGCAAAAGAGAAATTTCTTCTATA 2065



```
QY 1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2066 AGCATTCCTTTTATCCAGAAACACCTGTGAAGCACCAGATAGTTTCAAGACTTAAGCCA 2125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2126 GATTGGGTTTTGAGAGAGATAACATGGAAGAAATCACAATCCCTGCAAGCTATTCAA 2185
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1851 ThrIleLeuCysLysLeuGlnIle 1858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2186 ATGGTGATGATACGCTTGCCATT 2209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-10-172-118-1748
; Sequence 1748, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1748
; LENGTH: 2414
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_018136
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1748

Alignment Scores:
Pred. No.: 1,32e-32 Length: 2414
Score: 437.00 Matches: 210
Percent Similarity: 36.74% Conservative: 153
Best Local Similarity: 21.26% Mismatches: 331
Query Match: 4.59% Indels: 294
DB: 13 Gaps: 33

US-09-914-698-1 (1-1861) x US-10-172-118-1748 (1-2414)
QY 913 ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys 932
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AAATCTGTAGCAACTATTCAGGCTGCTTTAGAGCATGAAGTGTAGACAAAATTGAAA 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 933 LeuPheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyr 952
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 AATGTATCAGAGGAAAG---ATGGCAGCCATTGTTAAACCAATCTGCACCTGCTGTATC 118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr 971
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 AGAAGCAAAACTCAG---TATGAGCTGTTCAAAGTGAAGTGTATGATTCGAAGTGG 175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 972 TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla 991
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 TATAAAGCTTCGGCTTGTGTTTCACAGAACAGATATCATCTTCAAAGTAGGGCT 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 992 AlaIlePheLeuGlnArgIleTrpArgArgArgPheAlaLysLysLeuLeuAlaAla 1011
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 GCAGTAACAATTCAAAAGCTTTT---TGTAGAATGCTCAAGAAAACCTG----- 283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGln 1031
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 284 -----GAAACACAGAAAATGTGTCTGCCCTCAGGATTCAG 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1032 -----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg 1049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 TTCTTCTTCAGATGGCTGTGTAT-----CGAGAGAATTTGTTTCAG 358
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1050 GlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlu 1069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 CAGAAAAGAGCTGCTATCACTTACAGCATTTATTTAGGACG---TGG-----CAA 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArg 1089
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 ACCAGAAAACAGTTTTTACTATATAGAAAACGACGAGTGTGTTTTACAAAATCATCACA 466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1090 AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu 1109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 467 GCATTTCTGTCTGCAAAAACATCAAGACAAGTCATTTTACAGATCAGACAGTGTTATC 526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyr 1129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 527 ATTATTCAAGCTAGAAAGTAAGGATTTATACAGAAACGG----- 565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1130 SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg 1149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 ----- 565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln----- 1166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 -----AAGTTTCAGGAAATTAATAATAGCACCATATAAATTCAGGCTATCGG 613
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 AGGAGATATAGAGCCAAGAAA-----TATTTATGTAAAGTG 649
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 LysCysIleIleAsnValGln-----ArgArgTyrArgAlaThrLeuGlnMetArg 1203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 650 AAAGCTGCTGCAAGATTCAAGCCTGGTATAGATGTTGGAGAGCA----- 694
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1204 ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPhe 1223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 695 -----CACAAAGAATATCTAGCTGTATTAAGCTGTATAAATTTATCAAGGTGCTTC 748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLysValThr 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 749 TATACCAAACTAGAG-----AGAACACGGTTTTTGAATGTGACAGCATCAGCA 796
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ATTATCATTCAGAAAATCGAGAGCTATATCTTCCTGCAAGATAGCTCATGACACTTC 856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMet 1283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TTAATGATAAAGA----- 871
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 872 -----CATCGAGCTGCTTGTGTTTATCCAGCACAT 901
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 902 TATAGAGGA----- 910
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 -----TATAAAGGAAGGACGCTCTCTCTCGGCAG----- 940
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1344 TyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 941 -----AAATCTGCTGCTTTGATCATCAAAAATATATACGAGCAGGAGGCT 988
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1364 MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg 1383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 989 GGAAGCATGAAAGGATAAATATATTGAATTTAAAAAATCTACAGTTATCTTACACAGCA 1048
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

1384 ArgTyrArgGlyAsnLeuGluMetArgGlyGlnIleGluValTyrGlnLysGlnArgGln 1403  
1049 CTGGTGGCTGGTGGCTAGTACGAAAGAGATTTTAGAA-----CAGAGA 1093  
1404 AlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeuCysLysAla 1423  
1094 GCCAAATTCGACTT-----CTTCACCTTCAGTGCAGCT 1126  
1424 GlyTyrArgGlyLeuArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal 1443  
1127 GCATATTATCACCTGAAT----- 1144  
1444 GlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
1144 ----- 1144  
1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMetLysArg 1483  
1144 ----- 1144  
1484 ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
1145 -----GCTGTAGAAATTCAGAGGCTATATACTTTTACCTGGCTGTGAGAATGCTAAC 1198  
1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
1199 AAGCAGGTT-----AATTGAGTCTGTGATTGAGAGAGGTTTCAGCAAGA 1246  
1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
1247 TTACAAGAAAGAGATTTATTACAGAAATATCATAGCATC----- 1285  
1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
1286 -----AAAAAGATTGAGCATGAAGTCAAGAAATGTCAGGCCAGCGCA 1327  
1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
1328 AATAGCGCTGCATCAGTAATACAGAAAGCAGTCCGCAATTTCTCCGTAAGAACGAG 1387  
1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysVal 1603  
1388 GAAATTCACACTAGT-----GGAATCATTTAAA----- 1414  
1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGlnGlyLeuLeu 1623  
1415 ATTGAGGCATTATGGAGAGGCTATTCTTGAGAGAGAAATGATTGTACAAATTTAAA 1474  
1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGlnAlaLysAlaValAsnSerValArg 1643  
1475 GCTATACGACTAAGTCTTCAAGTTGTTAATAGGAGATTGAGAGAAACAAACTCTAC 1534  
1644 CysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663  
1535 AAAAGAACTGCACCTTGCACTTCACTTACCTTTTGACATATAAGCACCTTCTGCCATCTT 1594  
1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuMetTrpCys 1683  
1595 GAGGCTTTAAACACCTAGAGGTAGTACTAGA----- 1627  
1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703  
1628 -----TTGTCTCCACTTGTGTGAGACATAGGCCAGAGTCGAGCAATTTCTAAA 1678  
1704 AspLysGlnLeuIleGluArgCysSerArg----- 1713  
1679 ATATTGTGTTGATCCGAAGTTGTAATCGCAGTATTCCTTGATGGAAGTCATCATATAT 1738  
1714 -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
1739 GCTGTCAAGTCTTGTCTTAATGATATCAAGTATGAGAAACTACTTCAGCAGCTTATGAT 1798

1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
1799 GTAGAAAATTGTATAGATATCTATTGGAGCTTTTGCAGATATACCGAGAAAAGCCTGGT 1858  
1743 ---ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVal 1761  
1859 AATAAAGTTGCAGACAAAGCGGAGCAATTTTACAAAACCTTGTGTGTTGGCTATT 1918  
1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781  
1919 TTACTG-----AAGACAACAATAGAGCCTCTGATGTACGAGTAGGTCCAAAGTT 1969  
1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801  
1970 GTTGACCGTATTACAGTCTCTACAACTTACAGCTCATATAACATAAAATG-----AAT 2023  
1802 AlaArgLysProProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817  
2024 ACTGAAAGAATA-----CTTTACAGCAAGCAAGAGAAATTTCTTCTATA 2065  
1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830  
2066 AGCAATTCCTTTTATCCAGAAACACCTGTAAAGCACAGAAATAGTTTCAAGACTTAAGCCA 2125  
1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850  
2126 GATGGGTTTGAAGAGATACATGAGGAAGAAATCACAAATCCCTCGAAGCTATTCAA 2185  
1851 ThrIleLeuCysLysLeuGlnIle 1858  
2186 ATGGTGATGATACGCTTGGCAT 2209

## RESULT 9

US-10-188-832-81  
; Sequence 81, Application US/10188832  
; Publication No. US20040076955A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Bos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions  
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder  
; TITLE OF INVENTION: Cancer  
; FILE REFERENCE: 018501-002330US  
; CURRENT APPLICATION NUMBER: US/10/188,832  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US 60/302,814  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/310,099  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/343,705  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/372,246  
; PRIOR FILING DATE: 2002-04-12  
; NUMBER OF SEQ ID NOS: 207  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 81  
; LENGTH: 2414  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-188-832-81

## Alignment Scores:

Pred. No.: 1,32e-32 Length: 2414  
Score: 437.00 Matches: 210  
Percent Similarity: 35.74% Conservatives: 153  
Best Local Similarity: 21.26% Mismatches: 331  
Query Match: 4.59% Indels: 294  
DB: 17 Gaps: 33

US-09-914-698-1 (1-1861) x US-10-188-832-81 (1-2414)

Qy	1264	LeuHisLeuArgGluValThrIleLysLeuGlnArgGlyPheHisalaGlnLysSerMet	1288
Db	857	TTATGATAAAGA	871
Qy	1284	ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis	1303
Db	872	-----CATCGAGCTGCTTGTGTATCCCAAGCACAT	901
Qy	1304	TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla	1323
Db	902	TATACAGGA	910
Qy	1324	AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer	1343
Db	911	-----TATAAGGAAGCAGGTCTCTCTTCGGCAG	940
Qy	1344	TyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla	1363
Db	941	-----AAATCTGCTGCTTTGATCATACAAAAATATATACAGCAGCAGGAGGCT	988
Qy	1364	MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg	1383
Db	989	GGAAGCATGAAGCATATAAATATATTGAAATTAATAAAATCTACAGTTATCTTACAGCA	1048
Qy	1384	ArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGln	1403
Db	1049	CTGCTGGCTGGTTCGTAGTACGAAAAAGATTTTAGAA	1093
Qy	1404	AlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuCysLysAla	1423
Db	1094	GCCAAAATTCGACTT	1126
Qy	1424	GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal	1443
Db	1127	GCATATTATACCTGAAT	1144
Qy	1444	GlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln	1463
Db	1144	-----	1144
Qy	1464	AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMetLysArg	1483
Db	1144	-----	1144
Qy	1484	ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg	1503
Db	1145	-----GCTGTAGAAATTCAAAGAGCCTATAAACTTTACCTGGCTGTGAAGAATGCTAAC	1198
Qy	1504	GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn	1523
Db	1199	AAGCAGGTT	1246
Qy	1524	ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu	1543
Db	1247	TTACAGAAAGAGATTTTTCAGAAATATCATAGCATC	1285
Qy	1544	GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu	1563
Db	1286	-----AAAAGATTGAGCATGAAGTCAAGAAATGCTCTGAGCCACGCA	1327
Qy	1564	ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe	1583
Db	1328	AATAGGGCTGCATCAGTAATACAGAAAGCAGTGGCCCATTTTCTCTCCCTGTAAGAACGAC	1387
Qy	1584	GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysVal	1603
Db	1398	GAAAAATTCACACT	1414
Qy	1604	IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeu	1623
Db	1415	ATTGAGCATTTATGGAGGCTATCTCTGGAGGAGGAAAAATGATTGTACAAAATATAA	1474
Qy	1624	AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg	1643

```

Db 1475 GCTATACCACTAAGTCTTCAAGTTGTTAATAGGAGATTTCGAGAGAAAACAACTCTAC 1534
QY 1644 CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheLeuAlaSerAspAlaLeu 1663
Db 1535 AAAAGAACTGCACTTTCACCTTTCACCTTTCACCTTTCACCTTTCACCTTTCACCTTTC 1594
QY 1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1683
Db 1595 GAGGCTTAAACACCTAGAGGTAGTACTAGA----- 1627
QY 1684 SerGluPheMetSerThrPheCysTyrGlyLeuMetAlaGlnAlaLeuArgSerGluVal 1703
Db 1628 -----TTGTCTCCACTTGTGTGTCAGAACATGCGCCAGAGTGGAGCAATTTCTAAA 1678
QY 1704 AspLysGlnLeuLeuGluArgCysSerArg----- 1713
Db 1679 ATATTGTTTGAATCCGAGTTGTAATCGCAGTATCTTGTATGGAAGTCATCAGATAT 1738
QY 1714 -----IleLeuLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730
Db 1739 GCTGTGCAAGTCTTGTCTTAATGATCTAAGTATGAGAAACTACTTCAGCAGTTTATGAT 1798
QY 1731 GluGlyLeuValThrIleAlaGlnMetLeu----- 1742
Db 1799 GTAGAAAATTTGATAGATATCTATTGGAGCTTTTGCAGATATACCGAGAAAAGCCTGGT 1858
QY 1743 ---ArgTrpCysAspLysAspSerGluLeuPheAsnThrLeuCysThrLeuIleTrpVal 1761
Db 1859 AATPAGTTGTCAGACAAAGCGGAGCATTTTACAAAAAATTGTTGTTGGCTATT 1918
QY 1762 PheAlaHisCysProLysLysArgLysIleLeuHisAspTyrMetThrAsnProGluAla 1781
Db 1919 TTACTG-----AAGACAAATAAGAGCCTCTGATGACGAAGTAGTCCAAAGTT 1969
QY 1782 IleTyrMetValArgGluThrLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801
Db 1970 GTTGACCGTATTACAGTCTCTCAAACTTACAGCTCATATAACATAAAATG-----AAT 2023
QY 1802 AlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817
Db 2024 ACTGAAGATA-----CTTTACAAGCAAAAGAAATCTTCTAT 2065
QY 1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
Db 2066 AGCATTCCTTTATCCAGAAACACCTGTAGGACCAAGATAGTTTCAAGACTTAAGCCA 2125
QY 1831 AspPheGlyIleLeuArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
Db 2126 GATTGGGTTTGAAGAAGAGATAACATGGAAGAAATCACAAATCCCTGTCAAGCTATTCAA 2185
QY 1851 ThrIleLeuCysLysLeuGlnIle 1858
Db 2186 ATGGTGATGATACGCTTGGCAATT 2209

RESULT 10
US-10-437-963-53524/c
; Sequence 53524, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

```

```

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 53524
; LENGTH: 3069
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55716C.1
US-10-437-963-53524

Alignment Scores:
Pred. NO.: 1.34e-23 Length: 3069
Score: 348.50 Matches: 239
Percent Similarity: 36.04% Conservative: 187
Best Local Similarity: 20.22% Mismatches: 440
Query Match: 3.66% Indels: 317
DB: 1.7 Gaps: 42

US-09-914-698-1 (1-1861) x US-10-437-963-53524 (1-3069)

QY 419 AlaIleSerProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAla 438
Db 3051 GCGGTCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2995
QY 439 ProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLys 458
Db 2994 CCGCTAACCCCAATCTCGGCTCCAGAGGAACCCCTCCCT----- 2950
QY 459 LysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysVal 478
Db 2949 -----TCGGCCACCCCGACGACCCGG 2929
QY 479 ArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThr 498
Db 2928 CGCGCGCGGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2905
QY 499 CysIleAsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThr 518
Db 2904 -----CCTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2854
QY 519 MetTyrLeuAspGluGln-----AlaVal 526
Db 2853 CTGAGCTCGACAGTCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2794
QY 527 AspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAla--- 545
Db 2793 CGCGCTTCCCTCTCTCTCGCCACGTCGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2734
QY 546 -----AspLeuAspAla---Asp 550
Db 2733 TCGCGCTCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2674
QY 551 LeuAsnAsnLysIleAspVal-----GlyLysLeuPheAsnGlu 563
Db 2673 GTGAGGCAAGCGCGACGCGGTGATGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2614
QY 564 ValArgAsnLysGluLeuValAlaProThrLysGluGlnSerMetAsnTyrLeu 583
Db 2613 CACCGGGGTGTGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2554
QY 584 ThrLysTyrArgLeu---GluThrLeuArgLysAla-----AlaValGlu 597
Db 2553 ATGGCGGCTCTGTTGAGGATTCCTTGAGGAGGTGTGAGGAGGTGTGAGGAGGTGTGAGG 2494
QY 598 LeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyr----- 614
Db 2493 AGGATGGCAGCCACATGAGCAGGAGAGCGTGGAGAGGTGCTCTCATGATGTGCCCA 2434
QY 615 -----ValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeu 630
Db 2433 ATCTGCAAAACATTGATTAATGGAGGTTAAAGATGAAAGAACATTGCCCTTTGTTCT 2374
QY 631 AspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeu 650

```





Db 763 CTCAGCTAAGTATGCCATGCGATACTTTGCAACTGTGAGTGGTCTCTGCCAGTGAGGCCAA 822  
Qy 357 yGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAla----- 370  
Db 823 TGTGGAGAAAGGCTCTTGACCTCAACCCCATCATGGAGTCCATTGGAATGTAAGAAC 882  
Qy 371 -----GlnSerSerArgPheAsnLeuHis---GluValGly-----Ar 382  
Db 883 AACACGAGATGATATAGCAGCCGTTTGGGAAGTATATGAGATGGTTGGTTGATAGAG 942  
Qy 382 gLysSerValnysGlySerProValLysAsnProHisLysArgSerHisGluLeuSe 402  
Db 943 ATATCGAATCATTGGTGCATATAGCAATATATCTTTTAGAGAAATCCAGATGTATT 1002  
Qy 402 rPheSerAspAlaProSerAsnGluSerLeuTyArg----- 414  
Db 1003 CCAGGCAGAGAGAGAGAAATATCATCTTCTATCAGCTTTGTGCTCAGCAAGATT 1062  
Qy 415 -----AsnGluThrValAlaIl 420  
Db 1063 ACCTGAATTTAAATGCTACGATTAGGAAATGCAGATAACTTTAATTACACAAACAAGG 1122  
Qy 420 eSerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAl 440  
Db 1123 AGGCAGTCTCTGTGATTGAAGGATGGATGATGCAAGAGGATGGCACATCTAGGCAGGC 1182  
Qy 440 aAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysph 460  
Db 1183 CTGCACCTTGTAGGAATTAGTAATCTCAT-----CAATGGGNAITTT 1227  
Qy 460 eLys---LeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValAr 479  
Db 1228 CCGAATACTTCTGGCATCTCTCACTTAGGCAATGTTGATTATCATCCGAGATGCAGA 1287  
Qy 479 gAspThrSerIleGlnProSerValLysLeuTyArgSerGluLeuTyMetGlnThrCy 499  
Db 1288 CAGCTGCACAAATACCTCCCAAG-----CATGAACCTCTCTG 1323  
Qy 499 sIleAsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrWe 519  
Db 1324 CATC----- 1327  
Qy 519 tTyLeuAspGluGlnAlaValAspArgHisGln----- 530  
Db 1328 -TTCTGTGACCTCATGGTGTGGACTATGAGGAGATGTCTACTGGCTCTGCCATCGGAA 1386  
Qy 531 -----AlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAs 548  
Db 1387 ACTGGCTACTGCCACAGACATACATCAAGCCCATCTCCAAGCTGCGAGGCCACGAATGC 1446  
Qy 548 pAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsn----- 566  
Db 1447 CCGGATGCTTTGGCCAAAGCACATCTATGCCAAGCTCTTAACTGGATGTGATATAATGT 1506  
Qy 567 -LysGluLeuValValAlaProThrLysGluGlnGlnSerMetAsnTyLeuThrLysTy 586  
Db 1507 CAATCAGCTCTCCATCTGTCTGTCACACAGCACTCTTTATTATGGTGTCTAGACATTTA 1566  
Qy 586 rArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLe 606  
Db 1567 CGGATTTGAAACATTTTGAGATAAATAGTTTGAACAGTTT----- 1606  
Qy 606 uProCysSerLysValAlaValTyValAsnLysGlnAlaLeuArgIleArgSerAspAr 626  
Db 1607 ----TGCAATAAT-----TATGCAAAATGAA-----AACTACACCACAATT 1644  
Qy 626 gAsnLeuHisLeu-----AspValValMetGlnArgThrIleLeuGluLe 641  
Db 1645 CAATATGATGTCCTCAAAATTGGAGCAAGAAGATATATGAAGCAACAAATTCATGGAC 1704  
Qy 641 uLeuLeu-----CysPheAsnProLeuTyTrpLeuArgLeuGlyLe 654  
Db 1705 ACTCATAGATTTTATGATAATACAGCCTTGTATTAATCTTATAGAATCAAAACTAGGCAT 1764

Qy 654 uGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe 674  
Db 1765 TCTAGATTTACTGATGAGGAATGCAAGATGCTTAAAGGCACAGATGACACCTGGGCCCA 1824  
Qy 674 rThrPheIleLeuAsnArgLeuPheArgAsnLysCys-----GluGluGlnArgTy 691  
Db 1825 AAAATTTGACAAACACATTTG-----AACAAATGTGCACCTCTTTGAAAAGCCTCGTCT 1878  
Qy 691 rSer---LysAlaTyx----- 695  
Db 1879 ATCAACAAAGACTTTCATCTCATCCAACTTTTGTGCAAAAGTGAATACCACTGTGAAGG 1938  
Qy 696 -----ThrLeuThrGluGluTyzAlaGluThrIleLysLysLys 708  
Db 1939 ATTTCTCGAAAGAATAAAGACACCGTTTGTGAAGAACAAATTAAGATCTTTAAATCAAG 1998  
Qy 708 sSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIl 728  
Db 1999 CAAGTTTAAAG-----ATGCTACCAGAACTATTTCAAGATGATGAGAAGGCCAT 2046  
Qy 728 e-----VallyLysHisasnProCysLeuPh 736  
Db 2047 CAGTCCAACTTCAGCCACTCTCAGGGCGCACACCCCTCACAGAACTCTGCAAAAGCC 2106  
Qy 736 eValLysLysSerPro-----HisLysGluThrLysAspIleLeuLe 750  
Db 2107 CACCAAAAGGCAGACAGGCCCAAATGGCCAAAGAGCACAAAGAAAACAGTGGGGCATCAGTT 2166  
Qy 750 uArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLe 770  
Db 2167 CAGAACTCCCTGCACCTGCTT-----ATGGAGACACT 2199  
Qy 770 uGlyTyzValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyzAlaPheAsnAs 790  
Db 2200 CAATGCCACTACCTCTACTATGTGCGTGTATCAAGCTTAATGACTCAAGTTTCC 2256  
Qy 790 nLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLe 810  
Db 2257 ATTCAAGTTTGTAGAAGAGGGCAGTGCAGCAGCTGAGAGCATGTGGTGTCTCTG 2311  
Qy 810 uArgAspAspLeuThrArgGlnLeuArgValProAla----- 822  
Db 2312 -----GAAACCATCCGAATCAGTCGGCCGCGTTTCCCTCAGCGTGGAC 2355  
Qy 823 -----IleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGly-- 837  
Db 2356 TTACCAAGAAATTTTCAGCCCTACCGTGTCTTAATGAAGCAGAAAGATGTCTGAGTGA 2415  
Qy 838 -----AlaLeuGlyGluAlaAsnPhedI 845  
Db 2416 CAGAAAGCAACATGCAAGAATGTGTAGAGAAACTGATCTGGACAAGACAAATATCCA 2475  
Qy 845 nLeuGlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSe 865  
Db 2476 GTTTGGT-----AAGACAAAGATCTTTT 2499  
Qy 865 rLeuLeuTrpGlnLeuIleTyx-----LysPheArgSerProLysPheHisAlaAlaI 883  
Db 2500 CGTCCCGGTCAAGTGGCCTATCTAGAAAATTTAGAGCTGACAACTGAGAGCTGCTG 2559  
Qy 883 aThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGlnArgArgIl 903  
Db 2560 CATCCGGATCCAGAAG---ACCATCCGAGGGTGGCTG----- 2593  
Qy 903 eArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheAr 923  
Db 2594 -CTGAGAAGAAGTACCACGATCGCGAAGCAGCCATCACCATGCAGAGATACCTGG 2652  
Qy 923 gGlyHisGlnMetArgLysTyzValLysLeuPheLysThrGluArgThrGlnAlaAlaI 943  
Db 2653 GGGCTACCAAGGCCCGCATGCTATGCTTAAGTTTCTGCGC-----AGAACCAGACGCAAC 2706





Db 4687 CAAGAAGGAGGATGAGCAAAACCTGTTAAGAACCTGATTTGGAACCTGAAGCCACCGTGG 4746  
Qy 1549 YLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuArgCysSerMetProG 1569  
Db 4747 TGTAGAGTCAATTG-----ATCCAGG 4770  
Qy 1569 yPheGlnAlaArgAlaArgGlyPheMetAlaArg-----Ly 1581  
Db 4771 ATTACCGGCATATATCTCTGTTTCATGTGTGTTTCGACATGCTGACTACCTGAATGATCA 4830  
Qy 1581 sArgPheGlnAlaLeuMetThrProGluMetMetAspLeuLeuArgGlnLysArgAlaAl 1601  
Db 4831 GAAAGTAAAGTCGTGCTTAACATCAACAATTAACAGCATC-----AA 4872  
Qy 1601 aLysValIleGlnArg-----TyrTrpArgGlyTyrIle 1612  
Db 4873 AAAAGTATTGAGAAAGAGAGTGATGATTTGAAACCGTCTCTCTGCTCTTAACAC 4932  
Qy 1612 uileArg----- 1614  
Db 4933 ATGCCGATTTTGCACCTCTTGAACACGACAGTGGAGAGGCGCTTTATGAAGCACAA 4992  
Qy 1615 ----ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuAr 1633  
Db 4993 CACATCTCGCCAGATGAACACCTCCCTACCAATTT-----GACCTGGCTGATCG 5046  
Qy 1633 gGlnGluAlaLysAlaValAsnSerValArgCysLysVal--GlnGluAlaValArgPh 1652  
Db 5047 GCAG-----GTGCTGAGTGTGCTGGCCATTCAGATCTACCAGCAGCTGTCGCGGT 5097  
Qy 1652 eLeuArgGly-----ArgPheIleAlaSerAspAlaLeu----- 1663  
Db 5098 GTTAGAGAACATCCTTCAGCCCAATGTTCTCAGGCATGCTGGAACATGAACGATTC 5157  
Qy 1664 -----AlaValLeuSerGlnLeuAspAr 1671  
Db 5158 GGGCGTGTCTGGGTGAAGCCACAGGTTTGAGAAAGCGAACCTCCAGTATCGCGCATGA 5217  
Qy 1671 gLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCy 1691  
Db 5218 GGGCACTACACATGACATCCATCTCCGCGAGCTCACTCTCCACTCGTCATGCTG 5277  
Qy 1691 sTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuLeuArgCy 1711  
Db 5278 TCAGCATGGCATGACCTGAACTGATCAAGCAGGTGTGTCAGCAGAGTGTCTTACATCAT 5337  
Qy 1711 sSerArgIleIleLeuAsn----- 1717  
Db 5338 AGGGCCATCACCTGGAACACCTTCTCTCGGAAGACATGTGCTCTCTCGAGTAAAGG 5397  
Qy 1718 ----LeuAlaArgTyrAsnSerThrThrVal-----AsnThrPh 1729  
Db 5398 CATCAGATCAGGTACATGTCACTCACTGGAAGATGGCTGGTGCAAGATCTGAT 5457  
Qy 1729 eGlnGluGlyGly-----LeuValThrIleAlaGlnMetLeu----- 1741  
Db 5458 GAATAGTGGGCTAAAGAAACCTCGAACCTCTCATTCCAGGCTGCTCAACTTTGCAAGT 5517  
Qy 1742 -LeuArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVa 1761  
Db 5518 GAAAAAGAAACAGATGATGATGAGAA-----GCCATTGTCTAT-GTGCAATG 5567  
Qy 1761 lPheAlaHisCysPro----- 1766  
Db 5568 CTTTAACCTACTGCCAGATGTGAAAGTGTGAATTTGATATCTCCAGTTAATGAGTTG 5627  
Qy 1767 ----LysLysArgLysIleIleHisAspTyrMetThrAsn 1778  
Db 5628 AAGAAAGAGTCTCTGTGTCGTTCATTCGTACTATACAGAT 5667

RESULT 12

US-09-954-456-1601

; Sequence 1601, Application US/09954456

; Patent No. US20020115057A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can

; TITLE OF INVENTION: Sets

; FILE REFERENCE: 689290-76

; CURRENT APPLICATION NUMBER: US/09/954,456

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US/60/233,617

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: US/60/234,052

; PRIOR FILING DATE: 2000-09-20

; PRIOR APPLICATION NUMBER: US/60/234,923

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,134

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,637

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,638

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,711

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,720

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,840

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,863

; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 2276

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1601

; LENGTH: 14800

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-954-456-1601

Alignment Scores:

Pred. No.:	7,8e-20	Length:	14800
Score:	322.00	Matches:	363
Percent Similarity:	35.69%	Conservative:	289
Best Local Similarity:	19.87%	Mismatches:	682
Query Match:	3.38%	Indels:	493
DB:	9	Gaps:	72

US-09-914-698-1 (1-1861) x US-09-954-456-1601 (1-14800)

Qy	77	LysLysArgMetSerAlaAlaAlaProProSerLysGlnThr-----Trp	92
Db	3062	AGGAAGAGTCTCGTGCAGCGCTCATCTCCGAGCTCAAGACATCCGGGTGCAGCTGG	3121
Qy	93	ArgValThrAlaProSerArgProAlaAlaTrpAla-HisProProGlnAla-ProL	112
Db	3122	-----AGCCTGTGAGAGCGCGACCGTGCACCGCTCGCGTGGTGGTGGCGC	3163
Qy	112	euValGluLysAsnValTyrIleThrProGlnGluProValTyrIleSerProGlnP	132
Db	3164	TGGACAAAGAGCCGCGCAGCGAGTGTGCCAGCGC-----ATGCCCGAGCAGC	3211
Qy	132	roArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuValIleA	152
Db	3212	AGAAAGCCAGCGCAGAG-----	3228
Qy	152	spAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePhe-	171
Db	3229	-----GTGAGGGGGTGGGCAAGGGGTGCCCGGCTCTCTG	3265
Qy	172	-----ProAspAsnLeuAlaAlaTrpPro-----ThrProThrLeuLysG	185
Db	3266	CCGAGGGCGAGAGGTCTTTGGCCCTACACAGGCATCGCTCGCGGCCCGCCAGCTCGCT	3325
Qy	185	lyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuG	205
Db	3326	CGGAGCTGGAG-----CTGAGCTGGGCAAGCTGG	3355

QY	205	luAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnLys	225
Db	3356	AGCAGGTCGCGACGCTGCTCCATCTAC---CTGGAGAGCTCAAGACCATCAGCCTG	3412
QY	225	erLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysT	245
Db	3413	TGATCCCGCGCACCGAGGGGCCGAGGAGGTCTCAGGGCCACGAGGAGCAGCTCAAGG	3472
QY	245	hrThrThrIleValHisAlaThrHisThrArg-----AlaLeuAlaCysIleHisG	262
Db	3473	AGGCCAGCGCGTTCGGGCCACCCCTCCGGAGCTCGAGGCCACCAAGGCCTCTCTGAAGA	3532
QY	262	luGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeuLysA	282
Db	3533	AGCTGCGGGCCACGCGCGGAGCAGCAGCAGCCACGCTTCAGACGCCCTGCGGATCAGATGC	3592
QY	282	rgAspIleLeuValGlySerProLeu---ArgLysTyrSerGluSerMetLysAspL	301
Db	3593	GGGGGGCACAGAGGTGGGAGGCGACTGACAGCGCGCAGCGGAGCGCGGACTTGGAG	3652
QY	301	euSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnG	321
Db	3653	TGGAGCGCTGGCGGAGCGGGTCGCCAGGTGCTTGAGCGCTGGCAGCGCTGCTGGGCC	3712
QY	321	luMetLysIleArgSerIleGlu-----GlnAsnArgTyrTyrGlnG	335
Db	3713	AGACCGACTTGGCGACGCGGAGCTCGAGCAACTGGGCGCGCAGCTGCTGTACTACCGCG	3772
QY	335	lu-----GlnGlnIleGlnIleL	341
Db	3773	AGAGTCGAGACCCCTTGGGCGCTTGGTTCAGAGCGCCAGCGCGCGGAGGAGCATCC	3832
QY	341	ysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGln---	359
Db	3833	AGGCCATGCGCTGGCCACAGCCAGCTGTGCGGGAGCAGCTCGCGCAGGAGCAGGCC	3892
QY	360	-----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe	375
Db	3893	TGCTGGAGGAGATCGAGCGCCACGCGCAGAGGTCTGAGGAGTGCCAGAGTTTCGGAAC	3952
QY	376	-----AsnLeuHisGluValGlyArgLysSerValLysG	387
Db	3953	AGTACATCAAGCCATCAGAGCATGAACTTCAGCTCAGCTGAGTACGACGCG---CAGC	4009
QY	387	lySerProValLysAsnProHisLysArg-----	396
Db	4010	TTGAGCGGTGGCTCCCGGCCCAAGAACCCCAAGTCCAGTCCGGATCAGAGAGTGCA	4069
QY	397	-----ArgSerHisGluLeuSerPheSerAspAlaProSer---	408
Db	4070	TCGAGGAGTACGTGGACCTGCTGACGACTACAGCGAGCTGCACACACTGACGAGCCAGT	4129
QY	409	-----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProProL	424
Db	4130	ACATCAAGTTTCATCAGCGAGACTCTGGCGCATGGAGAGGAGGAGGAGCTGGCTGAGC	4189
QY	424	ysLysGlnArgValGluAspThrThrLeuProArgSerAlaIaProAlaAsnAlaSerA	444
Db	4190	AGCAGCGCGAGGAGGCGGCGGCTGGCGGAGTGGAGGCGCGCTGGAGAGGAGCAGC	4249
QY	444	laArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaG	464
Db	4250	GGCAGCTGGCGGAGGCGCACGCCAGGCAAGGCACAGCGCGGAGCGG---GAGCGCAAGG	4306
QY	464	InThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrSerIleG	484
Db	4307	AGCTGCAGCAGCGCATCGAGGAGGTGGTGGCGGAGGAGGCGCGCGCTGGAGCGCGC	4366
QY	484	lnProSerValLysLeuTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProA	503
Db	4367	AGCAGCAAGAGCGCAGCATTCAGGAGGAGCTGCGCAGCAGCTGCGCGCAGAGCTCGGAGCGG	4426

QY	503	spProPheAlaalaThrThrThrIleAspPropheLeuAlaSerThrMetTyrLeuAspG	523
Db	4427	AGATCCAGGCCAAGCCCGGCGAGCGCGGCTGCGCGCGCTGCGCATCGAGG	4186
QY	523	lu-----GlnAlaValAspArgHisGlnAlaAspPheL	534
Db	4487	AGGAGATCCGGTGGTGGCGCTGCAGTTGGAGGCGCACCGAGCGCGACGCTGGCGGGGCTG	4546
QY	534	ysLysTrpLeuAsnAlaValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn-	553
Db	4547	AGGGGGAGCTGCAGCACTCGTGCACGGCGCGAGGAGGTGAGGCAAAAGCGACAGG	4606
QY	554	--LysIleAspValGlyLysLeuPheAsnGluVal-	564
Db	4607	CGCAGGAGGCGCGAGCGCTTCGGGAGGCGAGGTGCAGGACGAGCAGCAGCGCGG	4666
QY	565	-----ArgAsnLysGluLeuValValAlaValAlaProThrLysLysGluG	577
Db	4667	AGGCGGAGGTGAGACTGGCTCGCGCTGAGGCGCGAGACCGAGGCGCGCGCGAGAGC	4726
QY	577	luGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThr-----L	591
Db	4727	AGCGGGCCCTCAGGCCCTGGAGGAGTGGCTGCAGCGGAGGCGCGAGCGCGCGCC	4786
QY	591	euArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys-	608
Db	4787	TGCGGCGCGCGAGGTGCAGCGCGGCGAGGTACAGTGGCCCTGGAGACGCGCGCAGC	4846
QY	609	-----SerLysValAlaValTyrValAsnLysGlnAlaLeuA	621
Db	4847	GCAGTCGAGGCGGAGCTGCAGAGCAACCGCGCTCTTCGCGGAGAGCAGCGCACAGC	4906
QY	621	rgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluL	641
Db	4907	TGGAGCGCTCCTCAGAGAGAACACCTGCTGTGGCACAGCTCGCGGAGGAGGCTGAG-	4965
QY	641	euLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluL	661
Db	4966	-----CGGCGGGCACAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4996
QY	661	ysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu-	679
Db	4997	GGGCGCGGAGGAGCAGCGGAG-----CTGGAGCGCTGGCAGCTCAAGGCCCA	5047
QY	679	snArgLeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrG	699
Db	5048	ACGAGGCGCTACGCTCGCGCTCAGCGGAGGAGGTGGCGCAGCAGCAGAGCTGGCGC	5107
QY	699	luGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProP	719
Db	5108	AGGCCGAGGCTGAG-----	5121
QY	719	heLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysL	739
Db	5122	-----AGCCAGAGGAG-----	5133
QY	739	ysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaA	759
Db	5133	-----	5133
QY	759	snIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHis-----A	777
Db	5134	-----GAGCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	5185
QY	777	rgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgA	797
Db	5186	GGCAGCGGAGCTGGCTGTGACAAAGACTGGAGAGCAGCGCGCGAGCTGGCGGAGCACCG	5245
QY	797	spGlyValArgLeuThrArgValValGluValIleLeuLeuArg-	812
Db	5246	CGCAGCAGCGCCTGGCGCGGACGACGAGTTGATCCGGCTGGCGGCGCGAGCGGACAGG	5305
QY	812	spAspLeuThrArgGlnLeuArgValProAlaIleSerArgTyrGlnArgIlePheAsnV	832

Db 5306 GGGAGCAGCGCGAGCTGTGGAGGAGGAGCTGGCCCGCTGACGGT----- 5355  
 Qy 832 allYsLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPhelGlnLeuGlyClyAspIleAlaA 852  
 Db 5356 -----GAGCGCGCTGCAGCCAGCAGCAACGCGAGAGCTGGAAGCCGAGCTGGCC- 5406  
 Qy 852 laGlnAspIleValAspGlyHisArgGluYsThrLeuSerLeuLeuTrpGlnLeuIleT 872  
 Db 5406 ----- 5406  
 Qy 872 yrLYsPheArgSerProLYsPheHisAlaAlaThrValLeuGlnYsTrpTrpArgA 892  
 Db 5407 --AAGTGGCGCGCGAG----- 5421  
 Qy 892 rgHisTrpLeuHisValValIleGlnArgArgIleArgHisLYsGlnLeuMetArgArgH 912  
 Db 5422 -----ATGGAGGTGTGTGTGCCCAAGCAAGCGGAGGCTGAGAGAGGTGCGCTCC- 5472  
 Qy 912 isArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLYsTrpValL 932  
 Db 5472 ----- 5472  
 Qy 932 ysLeuPheLYsThrGluArgThrGlnAlaAlaIleLeuGlnLYsPheThrArgArgT 952  
 Db 5473 -----ACCAGCGAGAAGTCCAAGCAGAGGCTGGAGCGCGAGCGCGCGTTCGCG 5524  
 Qy 952 yrLeuAlaGlnLYsGlnLeuTrpGlnSerLYsHisSerIleIleThrIleGlnArgTrpT 972  
 Db 5525 AGCTGGCGCGAGGAGCGCC----- 5544  
 Qy 972 rpArgAlaGlnLeuGlyArgGln---HisArgGlnArgPheValGluLeuArgGluA 991  
 Db 5545 --CGCTCGTGTCCCTGGCGAAGAGGCAAGCGCAGCG-----CAGCTGGCGGAG 5596  
 Qy 991 laAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLYsLYsLeuLeuAla 1011  
 Db 5597 AAGACGCGCGCGCGAGCGCGAGCGCGAGCGGTGCTTGGCGAGAAGCTGGCGCCA 5656  
 Qy 1011 laAlaGluThrAlaArgLeuGlnArgSerGlnLYsGlnAlaAlaAsnTrpIleG 1031  
 Db 5657 TCGCGAGCGCACCGCGCTC-----AAGACGAGCGCGAGATCGCGCTCAAG 5704  
 Qy 1031 lnMetGlnTrpArgThrLYsGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnA 1051  
 Db 5705 AAGAGCGCGGAGAACGAGCGCTCGCGCGCTGGCGAGACGAGCGCTTCCAGCGC 5764  
 Qy 1051 rgAspLeuIleMetPheValGlnArgArgMetArgSerLYsTrpSerMetLeuGluGlnA 1071  
 Db 5765 GCGCGCTG-----GAGGAGCAGCGCGCGCAACACAAAGCTGCACATCGAGGAGC 5812  
 Qy 1071 rgLYsGluPheGlnGlnLeuLYsArgAlaAlaIleAsnIleGlnArgTrpArgAlaL 1091  
 Db 5813 GC-----CTGCGCCAGCTGCGCAAGGATCGGACAGCGAGCTGGAGCGCGAGAGGGGC 5866  
 Qy 1091 ysLeu-----SerMetArgLYsCysAsnAlaAspTrpLeuAlaLeuArgS 1106  
 Db 5867 TGTGGAGCACCGCTGAGCGAGCGCGAGGTGGAGGAAGATCTGTGGCGCTGAAG 5926  
 Qy 1106 erSerValLeuLYsValGlnAlaTrpArgLYsAlaThrIleGlnMetArgIleAspArgA 1126  
 Db 5927 CGAGCTTCGAGAAGCGCGCGCT---GGCAAGCGGAGCTGAGCTGGAGCTGGGACGC- 5982  
 Qy 1126 snHisTrpTrpSerLeuArgLYsAsnVal---IleCysLeuGlnGlnArgLeuArgAlaI 1145  
 Db 5983 -----ATCCGCACACGCGGAGACACGCTGCGCAGCAGGAGGAGCGCGC 6028  
 Qy 1145 leMetLYsMetArgGluGlnArg-----GluAsnTrpLYsArgLeuArgAsnA 1161  
 Db 6029 ACCTGGAGCGCGGAGGAGCGGAGCTGGCGGAGGAGGAGGAGCGCGCGCTGTAGG 6088  
 Qy 1161 laSerIleLeuValGlnLYsArgTrpArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181

Db 6089 CTGAGGAGCGCGTGCAGAAAGACCTGGCGCGCGAGGAGAGCGCCGACGCGGAGAGG 6148  
 Qy 1181 laTrpLeu-----ArgThrArgLYsCysIleIleAsnValGlnArgArgTrpArgA 1198  
 Db 6149 CGGCGCTGGAGAACTGACGCGGTGAAGCAACGCTGGAGGAGAGCGCGCGCTGCGGG 6208  
 Qy 1198 laThrLeuGlnMetArgArgGluArgLYsAsnTrpLYsHisLeuGlnTrpThrLYsA 1218  
 Db 6209 AGCAGCGGAGCAGGAGTGGCGCGCGAGCTGCAGTGGCCGAGGAGCGCCGCAAGAGC 6268  
 Qy 1218 rgIleGlnIleLYsPheArgAlaLYsArgGluMetLYsLYsGlnArgAlaGluPheLeuG 1238  
 Db 6269 GGCTGCAGCGGAGAGAAAGCACAACGCTTCGCGGTGCAGCAGAAAGAGAGAGAGCTAC 6328  
 Qy 1238 lnLeuLYsLYsValThrLeuValGlnLYsArgArgAlaLeuLeuGlnMetArgL 1258  
 Db 6329 AG-----CAGACGCTGCAGAGGAGCAGAGCGTGTGGACGAGCTGCGCG 6373  
 Qy 1258 ysGluArgGlnGluTrpLeuHisLeuArgGluValThrIleLYsLeuGlnArgPheH 1278  
 Db 6374 CCGAGGCGGAG-----GCGCGCGCGCGCGCGCG 6400  
 Qy 1278 isAlaGlnLYsSerMetArgPheMetArgAlaLYsTrpArgGlyThrGlnAlaValS 1298  
 Db 6401 CTGAGGAGCGGAGGAGCGCGGTGAGCGCGGAGCTGAGCGCGCGAGGCC----- 6453  
 Qy 1298 erCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLYsArgGluArgAsnSerPheL 1318  
 Db 6454 -----CGCGCGCAGGTGGAAGGCGCG 6475  
 Qy 1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTrpArgAlaArgLeuAsnMetI 1338  
 Db 6476 AGCGCTGAAGCAGCTGGCA-----G 6496  
 Qy 1338 leLYsGlnLeuLYsSerTrpAlaGlnLeuLYsGlnAlaAlaIleThrIleGln----- 1355  
 Db 6497 AGGAGCAGGCACAGCGCGCTCAGGCACAGCGCGGTGCAGAGAGTTCGCGCAAGAGGAG 6556  
 Qy 1356 --ThrArgTrpArgAlaLYsLYsAlaMetGlnLYsGlnValLeuTrpGlnLYsGln- 1374  
 Db 6557 CCGAGCAGAGCGCGCGCGCGCGCACAGCGCGAGCGCGCTGCGCGAGAGCAGG 6616  
 Qy 1375 --ArgGluAlaIleIleLYsValGlnArgArgTrp---ArgGlyAsnLeuGluMetArgL 1393  
 Db 6617 CAGCTGCAGCGAGATGAGAGGAGATAAGAAATTCGCGCAGCAGCAGCTGCGCGCAAGAG 6676  
 Qy 1393 ysGlnIleGluValTrpGlnLYsGlnArgGlnAlaValIleArgLeuGlnLYsTrpTrpA 1413  
 Db 6677 CGCAGGTGGAG-----CAGGAGCTGACCAACTCGCGGTGCAG----- 6714  
 Qy 1413 rgSerIleArgAspMetArgLeuCysLYsAlaGlyTrpArgArgIleArgLeuSerSerL 1433  
 Db 6715 -----CTGGAGGAGACCGACACCAAG---AACCTGTCTGGAGC 6751  
 Qy 1433 euSerIleGlnArgLYsTrpArgAlaThrValGlnAlaArgArgGlnArg----- 1449  
 Db 6752 AGGAGCTGCAGCGCTGAAGCGGAGCGCACCGAGCGCGCACCGCAGCGCAGCCAGCTGG 6811  
 Qy 1450 ----GluIlePheLeuSerThrIleArgLYsValArgLeu-----MetGlnAlaPheI 1466  
 Db 6812 AGGAGAGAGCTTCTTCGTGGCGGTGAGAGTGGAGAGCTGAGCAAGCTCAAGGACCGCA 6871  
 Qy 1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnArgArgGluP 1479  
 Db 6872 TCGAGCTGAGAACCGCGCACTCATCTTCGCTGACAGGACAAATACGAGCGCTTCTGTC 6931  
 Qy 1479 heGluMetLYsArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaM 1499  
 Db 6932 AGGAGAGGCTGAGAGATGAAGCAGGTGTCGCGAGGAGCGCGCGGTGAGTGTGGCGG 6991  
 Qy 1499 etLeuLYsAlaArgGlnAspTrpGlnLeuIleGlnSerValIleLeuGlnArgL 1519  
 Db 6992 CCCAAGAGGCTCGCGACTGCGGAGCTGGCAGAGAGGAGAC-----CTGGCACAGCAGC 7045

Qy 1519 ysPheArgAlaAsnArgSerMetLys-----GlnAlaArgGlnGluPheValGlnL 1536  
Db 7046 GGGCGCTTGCAGAGAGATGCTCAGGAGAGATGCGAGCGGTGCAGGAGCCACGCGAC 7105  
Qy 1536 euArgThrIleAlaValHisLeuGln-----LysPheArgGlyLysA 1551  
Db 7106 TCAAGGCTGAGCGGAAGCTGTCAGCAGCAGAGAGAGAGCTTGCGCAGGAGCGCGCGC 7165  
Qy 1551 xgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheG 1571  
Db 7166 GGCTGCGAGGAGCAAGAGCAGATGCGCGAGAGCTGCGCGAGAGACCGCAGGCGCTTCC 7225  
Qy 1571 lnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGlu- 1590  
Db 7226 AG-----CGGACCTGGAGGCCGAGCGCGCAGCGCAGCTGGAGATGAGCGCTGAGG 7276  
Qy 1591 -----MetMetAspLeuIleArg----- 1596  
Db 7277 CTGAGCGCCTCAAGCTCGTGTGCGCGAGATGAGCGAGCCAGCGCGCGCTGAGGAGG 7336  
Qy 1597 -----GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrL 1612  
Db 7337 ACGCCAGCGCTTCCGAAGAGCGCGAGGAGATCGGTGAGAAGCTGCACCGCGCAGG 7396  
Qy 1612 eu-----IleArgArgGlnLysH 1619  
Db 7397 TCGCCACCCAGGAGAGTGCCTGTGTCAGACACTGAGATCCAGCGACAGCAGAGTG 7456  
Qy 1619 isGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaV 1639  
Db 7457 ACCATGATCGCAGCGCTGCGGAGGCCATCGTGAGCTGGAGCGTGAGAAGGAG--- 7512  
Qy 1639 alaSerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleA 1659  
Db 7513 -----AAGCTCAACAGGAGGCCAAACTGCTGCAG-----CTCA 7546  
Qy 1659 laSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrVal-PropHis 1678  
Db 7547 AGCTGAGGAGATGCAGCGTGCAGCAGGAGCAGCTGCTGCAGGAGACCGAGCC--- 7603  
Qy 1679 LeuLeuMetTrpCysSerGluPheMetSer-----ThrPheCysTyrGlyIleMet 1695  
Db 7604 -----TGCAGCAAGCTTCTCTCTGAAAAGGACAGCTGCTACAGCGGAGC 7651  
Qy 1696 AlaGlnAlaIleArg 1700  
Db 7652 GCTTCATCGAGG 7666

## RESULT 13

US-10-269-909-61  
; Sequence 61, Application US/10269909  
; Publication No. US20030180747A1  
; GENERAL INFORMATION:  
; APPLICANT: HRUBAN, RALPH H.  
; APPLICANT: ARGANI, PEDRAM  
; APPLICANT: IACOBUIO-DONAHUE, CHRISTINE  
; APPLICANT: MAITRA, ANIRBAN  
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES  
; FILE REFERENCE: 58303 (71699)  
; CURRENT APPLICATION NUMBER: US/10/269,909  
; PRIOR FILING DATE: 2003-10-11  
; PRIOR APPLICATION NUMBER: 60/328,609  
; PRIOR FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 60/332,754  
; PRIOR FILING DATE: 2001-11-19  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 61  
; LENGTH: 14800  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-269-909-61

Alignment Scores:  
Pred. No.: 7.8e-20 Length: 14800  
Score: 322.00 Matches: 363  
Percent Similarity: 35.69% Conservative: 289  
Best Local Similarity: 19.87% Mismatches: 682  
Query Match: 3.38% Indels: 493  
DB: 72 Gaps: 15  
US-09-914-698-1 (1-1861) x US-10-269-909-61 (1-14800)  
Qy 77 LysLysArgMetSerAlaAlaAlaProProSerLysGlnThr-----Tyr 92  
Db 3062 AGGAAGAGTCTCGTGCAGCGTGCATCTCCGAGCTCAAGACATCCGGCTGCAGCTGG 3121  
Qy 93 ArgValThrAlaProSerArgProAlaAlaTrpAla-HisProProGlnAla-ProL 112  
Db 3122 -----AGGCTGTGAGACGCGCAGCTGCACCGCTGCGGCTGCGGC 3163  
Qy 112 euValGluLysAsnValTyrLysThrProGlnGluProValTyrLysSerProGlnP 132  
Db 3164 TGGACAAGAGCGCGCAGGAGTGTCCAGCGC-----ATCGCGAGCAGC 3211  
Qy 132 roArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleA 152  
Db 3212 AGAAGGCACAGGCGAG----- 3228  
Qy 152 spAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePhe- 171  
Db 3229 -----GTGAGGGGCTGGGCAAGGGGCTGCCCGGCTCTCTG 3265  
Qy 172 -----ProAspAsnLeuAlaAlaTrpPro-----ThrProThrLeuLysG 185  
Db 3266 CCGAGCGCGAGAGGCTTGGCCCTACGAGCCATCGCTGCGGCCCCACCGCTGCGCT 3325  
Qy 185 lyAsnValLysSerCysAlaAsnAspMetArgProArgIleThrProAspAspLeuG 205  
Db 3326 CGGAGCTGGAG-----CTGACGCTGGGCAAGCTGG 3355  
Qy 205 luAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnLys 225  
Db 3356 AGCAGGTCCGACAGCTGTCTGCCATCTAC---CTGGAGAAGCTCAAGACCATCAGCCTG 3412  
Qy 225 erLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysT 245  
Db 3413 TGAITCCGGCCACGCGAGGGGCGGAGAGTGTCTCAGGGCCACGAGGACGAGCTCAAGG 3472  
Qy 245 hrThrThrIleValHisAlaThrHisThrArg-----AlaLeuAlaCysLysHisG 262  
Db 3473 AGGCCCGAGCGGTGCGCGCCACCTCCCGAGGCTCGAGGCCACCAAGGCTCTCTGAGA 3532  
Qy 262 luGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeuLysA 282  
Db 3533 AGCTGCGGGGCCAGGCGGAGCAGCAGCCACGTTCCAGCGCCTCGCGGATGAGCTGC 3592  
Qy 282 xgAspIleLysLeuValGlySerProLeu---ArgLysTyrSerGluSerMetLysAspL 301  
Db 3593 GGGGGGCACAGGAGGTGGGGAGCGAGTGCAGCAGCGGCGCAGCGGAGCGCGGACTTGGAGG 3652  
Qy 301 euSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnG 321  
Db 3653 TGGAGCGCTGGCGGAGCGGTGCTCCCGAGTGTCTGAGCGCTGCGAGGCTGTCTGGCCC 3712  
Qy 321 luMetLysIleArgSerIleGlu-----GlnAsnArgTyrTyrGlnG 335  
Db 3713 AGACCGACTTGGCGCAGCGAGCTCGACCACTGGGCGCGCAGCTGCGTACTACCGCG 3772  
Qy 335 lu-----GlnGlnIleGlnIle 341  
Db 3773 AGAGTCAGACCCCTTGGCGCTGCTGCTGAGCACGCCAGCGCGCGGAGCAGATCC 3832  
Qy 341 ysAlaLysAspLeuAsnSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 359

Db 3833 AGGCATCGCGTGGCCGACACCGCAGGCTGTGCGGAGCAGCTGCGGAGGAGCGCC 3892  
 Qy 360 -----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe----- 375  
 Db 3893 TGTGGAGGAGATCGAGCGCCAGCGCGAGAGAGGTCTGAGAGGTGCGAGAGGTTTGGGAAC 3952  
 Qy 376 -----AsnLeuHisGluValGlyArgLysSerValLysG 387  
 Db 3953 AGTACATCAACGCCATCAAGGACTATGAATCCAGCTCGTGGTACGCTACAGGCG--CAGC 4009  
 Qy 387 lySerProValLysAsnProHisLysArg----- 396  
 Db 4010 TTGAGCGGTGCGCTCCCGGCCAAGAAAGCCAGGTCAGTCCAGTCCAGGAGTGTCA 4069  
 Qy 397 -----ArgSerHisGluLeuSerPheSerAspAlaProSer---- 408  
 Db 4070 TCCAGGAGTACGTGGAACCTGCGTACGCACCTACAGGAGCTACACACCTGAGGCGCAGT 4129  
 Qy 409 -----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProL 424  
 Db 4130 ACATCAAGTTTCATCAGCGAGACTCTGCGCGCATGAGGAGGAGGAGGCTGGCTGAGC 4189  
 Qy 424 yLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerA 444  
 Db 4190 AGCAGCGGCGAGAGGAGCGCGAGCGGCTGGCGGAGTGGAGGCGCGCTGGAGAACGAGC 4249  
 Qy 444 laArgSerSerSerAlaHisAlaIleProHisAlaGlnSerLysLysPheLysLeuAlaG 464  
 Db 4250 GGCAGCTGGCGAGCGCGCCAGGCGAAGGCGAAGGCGAGGCGG--GAGGCGAAGG 4306  
 Qy 464 lnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrSerIleG 484  
 Db 4307 AGCTGCACGCGCATGCGAGGAGGAGTGTGCGCGGAGGAGGCGCGCTGGAGCGCGC 4366  
 Qy 484 lnProSerValLysLeuTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProA 503  
 Db 4367 AGCAGAGAGCGCAGCATTCAGGAGGAGCTGCAGCAGCTGCGGAGAGCTGCGGAGCGCG 4426  
 Qy 503 spProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspG 523  
 Db 4427 AGATCCAGGCCAAGGCCCGGAGGAGGCGGCTGAGGCGAGCGCGCTGCGCATCGAGG 4486  
 Qy 523 lu-----GlnAlaValAspArgHisGlnAlaAspPheL 534  
 Db 4487 AGGAGATCGCGTGTGCGCTGCGTGAAGCGCGAGACCGAGCGCGCTGCGCGGGCTG 4546  
 Qy 534 yLysTyrPheAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn- 553  
 Db 4547 AGGGGAGCTGCAGGCACTGCGTGCACGGCGGAGGAGGCTGAGGCGCAAAAGCGACAGG 4606  
 Qy 554 --LysIleAspValGlyLysLeuPheAsnGluVal----- 564  
 Db 4607 GCAGAGGAGGCCGAGCGCTTGGGAGGCGCGGTGCGAGCAGAGCGACGCGTAAGCGGC 4666  
 Qy 565 -----ArgAsnLysGluLeuValValAlaProThrLysGluG 577  
 Db 4667 AGGCGGAGTGGAGCTGCGCTGCGCGTGAAGCGCGAGACCGAGCGCGCGCGCGAGAGC 4726  
 Qy 577 luGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThr-----L 591  
 Db 4727 AGCGGCGCTGAGGCGCTTGGAGGAGCTGCGGCTGCGAGCGGAGGCGGAGCGCGCGCC 4786  
 Qy 591 euArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys----- 608  
 Db 4787 TGGCGAGGCCGAGGTGGAGCGAGCGCGCGCAGGTACAGTGGCCCTGGAGACGGCGCAGC 4846  
 Qy 609 -----SerLysValAlaValTyrValAsnLysGlnAlaLeuA 621  
 Db 4847 GCAGTGCAGAGCGGAGCTGCAGAGCAACGCGCTCTTCGCCGAGAGACGCGCACAGC 4906  
 Qy 621 rgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 641  
 Db 4907 TGGAGCGCTCCCTGCGAGGAGAACACGTGGTGTGGCAGCTGCGGAGGAGGAGGTGAG- 4965

Qy 641 euLeuLeuCysPheAsnProLeuTyrPheArgLeuGlyLeuGluValValPheGlyGluL 661  
 Db 4966 -----CGGCGGCACAGCAGCAGCGCCGCGCGAGC 4996  
 Qy 661 ysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu-----A 679  
 Db 4997 GGGCGCGAGGAGGAGCGAGCGGAG-----CTGGAGCGCTGGCAGCTCAAGGCCA 5047  
 Qy 679 snArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrG 699  
 Db 5048 ACGAGGCGCTACGCGTGGCTGCGCGGAGGAGGTGCGCAGCAGAGAGCTTGGCGC 5107  
 Qy 699 luGluTyrAlaGlnThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProP 719  
 Db 5108 AGGCGGAGCTGAG----- 5121  
 Qy 719 heLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysL 739  
 Db 5122 -----AAGCAGAAGGAG----- 5133  
 Qy 739 ysSerProHisLysGlnThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaA 759  
 Db 5133 ----- 5133  
 Qy 759 snIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHis-----A 777  
 Db 5134 -----GAGCGGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 5185  
 Qy 777 rgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgA 797  
 Db 5186 GGCAGCGGAGCTGGCTGAACAAGAGCTGGAGAGCAGCAGCGGAGCTGCGGAGGAGCGC 5245  
 Qy 797 spGlyValArgLeuThrArgValValGluValIleLeuLeuArg-----A 812  
 Db 5246 CCGAGCAGCGCTGCGCGCGAGCGAGGAGTGTATCCGCTGCGGCGCGAGCAGCGAGG 5305  
 Qy 812 spAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnV 832  
 Db 5306 GGGAGCAGCAGCGCGAGCTGTGGAGGAGGAGCTGCGCGCGCGTGCAGCGT----- 5355  
 Qy 832 allYsLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaA 852  
 Db 5356 -----GAGCGCGCTGCAGCCAGCAGAACCGCAGAGCTGGAACCGCAGCTGGCC- 5406  
 Qy 852 laGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeuIleT 872  
 Db 5406 ----- 5406  
 Qy 872 yrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGlnLysTyrTrpArgA 892  
 Db 5407 --AAGGTGCGGCGCGAG----- 5421  
 Qy 892 rgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGlnLeuMetArgArgH 912  
 Db 5422 -----ATGAGGTGCTGCTGCACCAAGCGGAGGCTGAGAGAGTGCCTGCTCC- 5472  
 Qy 912 isArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValL 932  
 Db 5472 ----- 5472  
 Qy 932 ysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgT 952  
 Db 5473 -----ACCAGCGAGAAGTCCAAGCAGAGAGCTGGAGCGCGAGCGCGGTTCCGCG 5524  
 Qy 952 yrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrpT 972  
 Db 5525 AGCTGCGCGAGGAGCGCGC----- 5544  
 Qy 972 trpArgAlaGlnGlnLeuGlyArgGln---HisArgGlnArgPheValGluLeuArgGluA 991  
 Db 5545 --CGCTGCGTCCCTGCGGGAAGAGCGCAAGCGCGAGCGG-----CAGCTGGCGAGG 5596



Qy 991 laAlaIlePheLeuGlnArgIleTrrpArgArgLeuPheAlaLysLysLeuLeuAla 1011  
 Db |||||  
 Qy 5997 AAGACGGCGCGGAGCGCGCGCGGAGCGGGTCTTGGCGAGAGCTGGCGCGCA 5656  
 Db |||||  
 Qy 1011 laAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnAlaAlaSerTyrIleG 1031  
 Db |||||  
 Qy 5657 TCGCGAGCGCGCGCGCTC-----AAGACGGAGCGGAGATCGCGCTCAAGG 5704  
 Db |||||  
 Qy 1031 lnMetGlnTrrpArgTrrpGlnLeuGlnArgIleGlnArgHisGluPheLeuArgGlnA 1051  
 Db |||||  
 Qy 5705 AGAAGGAGCGGAGAAACAGCGCTGGCGCGCTGGCGGAGGAGAGCGCTTCACGCGG 5764  
 Db |||||  
 Qy 1051 rgAspLeuIleMetPheValGlnArgMetArgSerLysTrrpSerMetLeuGlnGlnA 1071  
 Db |||||  
 Qy 5765 GCGCGCTG-----GAGGAGCGCGCGCAACAGCTGACATCGAGGAGC 5812  
 Db |||||  
 Qy 1071 rgLysGluPheGlnGlnLeuLysArgAlaIleAsnIleGlnArgTrrpArgAlaL 1091  
 Db |||||  
 Qy 5813 GC-----CTGGCCAGCTGGCAAGCATCGGACAGCGAGCTGGAGCGGCAAGGCGG 5866  
 Db |||||  
 Qy 1091 ysLeu-----SerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgS 1106  
 Db |||||  
 Qy 5867 TGGTGGAGGACACCTGAGGCGCGCGGAGGTGGAGAGAGATCTCGCGCTGAAGG 5926  
 Db |||||  
 Qy 1106 erSerValLeuLysValGlnAlaTrrpArgLysAlaThrIleGlnMetArgIleAspArgA 1126  
 Db |||||  
 Qy 5927 CGAGCTTCGAGAGCGCGCT---GGCAAGCGGAGCTGGAGCTGGAGCTGGAGCGC- 5982  
 Db |||||  
 Qy 1126 snHisTyrTrrpSerLeuArgLysAsnVal---IleCysLeuGlnGlnArgLeuArgAlaI 1145  
 Db |||||  
 Qy 5983 -----ATCCGACGAAACGCGGAGGACACGCTGCGCAGCAAGGAGGAGCGCG 6028  
 Db |||||  
 Qy 1145 leMetLysMetArgGlnGlnArg-----GluAsnTrrpLeuArgLeuArgAsnA 1161  
 Db |||||  
 Qy 6029 AGCTGGAGCGCGGAGGACGCGCAGCTGGCGCGGAGGAGGAGCGCGCGCGCTGAGG 6088  
 Db |||||  
 Qy 1161 laSerIleLeuValGlnLysArgTrrpArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181  
 Db |||||  
 Qy 6089 CTGAGGAGCGCTGCAGAGAGCTGGCGCGCGGAGGAGGCGCGCACGCGCGGAGG 6148  
 Db |||||  
 Qy 1181 laTrrpLeu-----ArgThrArgLysCysIleIleAsnValGlnArgTrrpArgA 1198  
 Db |||||  
 Qy 6149 CGCGCTGGAGGAAGTGCAGCGGCTGAAAGCAACGTTGGAGGAGCGCGCGCTGCGGG 6208  
 Db |||||  
 Qy 1198 laThrLeuGlnMetArgArgGluArgLysAsnTrrpLeuHisLeuGlnThrThrLysA 1218  
 Db |||||  
 Qy 6209 AGCGAGCGGAGGAGTTCGGCGCGGAGCTGCAGCTGGCCCGGAGGCGCGCGCGAGG 6268  
 Db |||||  
 Qy 1218 rgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuG 1238  
 Db |||||  
 Qy 6269 GGCTGCAGCGGAGAGAGACACCGCTTCGGGTGCAGAGAGAGGAGGAGGAGCTAC 6328  
 Db |||||  
 Qy 1238 lnLeuLysLysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgL 1258  
 Db |||||  
 Qy 6329 AG-----CAGACGCTGCAGCAGGAGCAGCGCTGCTGGACCACTGCTGCGG 6373  
 Db |||||  
 Qy 1258 ysGluArgGlnGluTrrpLeuHisLeuArgGluValThrIleLysLeuGlnArgPheH 1278  
 Db |||||  
 Qy 6374 GCGAGCGGAG-----GCGGCGCGCGCGCGG 6400  
 Db |||||  
 Qy 1278 isAlaGlnLysSerMetArgPheMetArgAlaLysTrrpArgGlyThrGlnAlaAlaVal 1298  
 Db |||||  
 Qy 6401 CTGAGGAGCGGAGGAGCGCGGTGCAGCGGAGCGCTGAGGCGCGCGCGAGCC- 6453  
 Db |||||  
 Qy 1298 erCysLeuGlnMetHisTrrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheL 1318  
 Db |||||  
 Qy 6454 -----CGCGCGCAGCTGGAAGAGCGCG 6475  
 Db |||||  
 Qy 1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnArgTrrpArgAlaArgLeuAsnMetI 1338  
 Db |||||  
 Qy 6476 AGCGGCTGAAGCAGCTCGCA-----G 6496  
 Db |||||  
 Qy 1338 leiLysGlnLeuLysSerTrrpAlaGlnLeuLysGlnAlaAlaIleThrIleGln----- 1355  
 Db |||||

Db 6497 AGGAGCAGGACAGCGCGCGCTCAGCAGCAGGCGCGCTGCAGAGAGCTGCGCAAGGAGG 6556  
 Qy |||||  
 Qy 1356 --ThrArgTrrpArgAlaLysLysAlaMetGlnLysGlnValValLeuTrrpGlnLysGln- 1374  
 Db |||||  
 Qy 6557 CCGAGCAAGAGGCGCGCGCGGACAGCGCGGAGCAGCGCGCTTGGCGGAGAGCAGG 6616  
 Db |||||  
 Qy 1375 --ArgGluAlaIleIleLysValGlnArgArgTrrp---ArgGlyAsnLeuGlnMetArgL 1393  
 Db |||||  
 Qy 6617 CAGCTGACGCGGAGATGAGAAATTCGCGGAGCAGACGCTGCGGCGAGAAGG 6676  
 Db |||||  
 Qy 1393 ysGlnIleGluValTrrpGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTrrpA 1413  
 Db |||||  
 Qy 6677 GCGAGGTGGAG-----CAGGAGCTGACACACCTGCGGCTGAG----- 6714  
 Db |||||  
 Qy 1413 rgSerIleArgAspMetArgLysAlaGlyTrrpArgArgIleArgLeuSerSerL 1433  
 Db |||||  
 Qy 6715 -----CTGGAGGAGACCGACACCAGAG---AACCTGCTGGAGC 6751  
 Db |||||  
 Qy 1433 euSerIleGlnArgLysTrrpArgAlaThrValGlnAlaArgArgGlnArg----- 1449  
 Db |||||  
 Qy 6752 AGGAGCTGACGCGCTGAGCGCGAGCCACGAGGCGCGCACGCGCAGCGCAGCGTGG 6811  
 Db |||||  
 Qy 1450 ----GluIlePheLeuSerThrIleArgLysValArgLeu-----MetGlnAlaPheI 1466  
 Db |||||  
 Qy 6812 AGGAGGAGCTCTTCTCGTGGCGCTGCAGATGAGGAGCTGAGCAAGCTCAAGGCGCGCA 6871  
 Db |||||  
 Qy 1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnArgArgGluP 1479  
 Db |||||  
 Qy 6872 TCGAGGCTGAGAACCGCGGCACTCATCTGCTGACAGGACATACCGAGGCTTCCTGC 6931  
 Db |||||  
 Qy 1479 heGluMetLysArgArgAlaValIleGlnArgArgPheArgAlaArgCysAlaM 1499  
 Db |||||  
 Qy 6932 AGGAGGAGGCTGAGAGATGAAGCAGGTGGCGGAGGAGCGCGCGCTGAGTGGCGG 6991  
 Db |||||  
 Qy 1499 etLeuLysAlaArgGlnAspTrrpGlnLeuIleGlnSerValIleLeuValGlnArgL 1519  
 Db |||||  
 Qy 6992 CCCAGAGGCTGCGGCACTGCGGAGCTGCGCAGCAGAGGAGGAGC-----CTGGCAGCAGC 7045  
 Db |||||  
 Qy 1519 ysPheArgAlaAsnArgSerMetLys-----GlnAlaArgGlnGluPheValGlnL 1536  
 Db |||||  
 Qy 7046 GGGCTTGGCAGAGAGATGCTCAGGAGAGAGATGCGGCGGTGCAGAGCGCGCACGCGAC 7105  
 Db |||||  
 Qy 1536 euArgThrIleAlaValHisLeuGlnGln-----LysPheArgGlyLysA 1551  
 Db |||||  
 Qy 7106 TCAAGGCTGAGCGGAACTGCTCAGCAGCAGAGAGGAGCTTGGCGCAGGAGCGCGCGG 7165  
 Db |||||  
 Qy 1551 rgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheG 1571  
 Db |||||  
 Qy 7166 GGCTGCAGGAGGACAGGAGCAGATGCGGAGCAGCTGGCGGAGGAGACCGCAGGCGCTTC 7225  
 Db |||||  
 Qy 1571 lnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGlu- 1590  
 Db |||||  
 Qy 7226 AG-----CGGACGCTGAGCGCGGAGCGCGCAGCTGGAGATGAGCGCTGAGG 7276  
 Db |||||  
 Qy 1591 -----MetMetAspLeuIleArg----- 1596  
 Db |||||  
 Qy 7277 CTGAGCGCTCAAGCTGCTGGCGGAGATGAGCGGAGCGCGCGCGCTGAGGAGG 7336  
 Db |||||  
 Qy 1597 -----GlnLysArgAlaAlaLysValIleGlnArgTrrpArgGlyTrrL 1612  
 Db |||||  
 Qy 7337 ACGCCAGCGCTTCGCGAAGAGCGGAGGAGATCGGTGAGAGCTGCACCGCAGCGAGC 7396  
 Db |||||  
 Qy 1612 eu-----IleArgArgArgGlnLysH 1619  
 Db |||||  
 Qy 7397 TCGCCACCCAGGAGAGGTGACCTGCTGTCAGACACTGGAGATCCAGCGCACAGCAGTG 7456  
 Db |||||  
 Qy 1619 isGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlav 1639  
 Db |||||  
 Qy 7457 ACCATGATCCGAGCGCTGCGGAGGCGCATCGCTGAGCTGGAGCGCTGAGAGGAG--- 7512  
 Db |||||  
 Qy 1639 alAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleA 1659  
 Db |||||



Db 7513 -----AAGCTCCACAGCAGGCGCAAACTGCTGCAG-----CTCA 7546  
 Qy 1659 laSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrVal-ProHis 1678  
 Db 7547 AGTCTGAGGATGACAGCGTGCAGGAGGAGCAGCTGCTCAGGAGCAGGAGGCC--- 7603  
 Qy 1679 LeuLeuMetTrpCysSerGluPheMetSer-----ThrPheCysTyrGlyleMet 1695  
 Db 7604 -----TGCAGCAAGAGCTTCCTCTCTGAAAGGACAGCTGCTACAGCGGGAGC 7651  
 Qy 1696 AlaGlnAlaIleArg 1700  
 Db 7652 GCTTCATCGACGAGG 7666

RESULT 14

US-10-717-597-183  
 ; Sequence 183, Application US/10717597  
 ; Publication No. US20040110221A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wyeth  
 ; APPLICANT: Burczynski, Michael E.  
 ; APPLICANT: Twine, Natalie C.  
 ; APPLICANT: Dorner, Andrew J.  
 ; APPLICANT: Trepicchio, William L.  
 ; APPLICANT: Slonim, Donna K.  
 ; APPLICANT: Stover, Jennifer A.  
 ; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS  
 ; FILE REFERENCE: AM101080L  
 ; CURRENT APPLICATION NUMBER: US/10/717,597  
 ; CURRENT FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: US 60/459,782  
 ; PRIOR FILING DATE: 2003-04-03  
 ; PRIOR APPLICATION NUMBER: US 60/427,982  
 ; PRIOR FILING DATE: 2002-11-21  
 ; NUMBER OF SEQ ID NOS: 4904  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 183  
 ; LENGTH: 14800  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-717-597-183

Alignment Scores:

Pred. No.:	7,8e-20	Length:	14800
Score:	322.00	Matches:	363
Percent Similarity:	35.69%	Conservative:	289
Best local Similarity:	19.87%	Mismatches:	682
Query Match:	3.38%	Indels:	493
DB:	17	Gaps:	72

US-09-914-698-1 (1-1861) x US-10-717-597-183 (1-14800)

Qy 77 LysLysArgMetSerAlaAlaAlaProProSerLysGlnThr-----Trp 92  
 Db 3062 AGGAAGAGTCTGCTGCCAGCGCTGCATCTCCGAGCTCAAAGACATCCGGCTGCAGCTGG 3121  
 Qy 93 ArgValThrAlaProSerArgProAlaAlaTrpAla-HisProProGlnAla-ProL 112  
 Db 3122 -----AGGCTGTGACACGCGCACCGTGCACCGCTGCCGCTGCCGCG 3163  
 Qy 112 euValGluLysAsnValTyrLysThrProGlnGluProValTyrLysSerProGlnP 132  
 Db 3164 TGGACAAAGACCGCGCAGCGGAGTGTGCCAGCGC-----ATCGCCGAGCAGC 3211  
 Qy 132 roArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuAspValIleA 152  
 Db 3212 AGAAGGCAAGGACAGAG----- 3228  
 Qy 152 spAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePhe- 171  
 Db 3229 -----GTGAGGGGCTGGGCAAGGGGCTGCCCGGCTCTCTG 3265  
 Qy 172 -----ProAspAsnLeuAlaAlaTrpPro-----ThrProThrLeuLysG 185

Db 3266 CCGAGGCGGAGAGGTTCTTGCCCTTACAGAGCCATGCGCTGCGGCCCGCCCGCTGCGCT 3325  
 Qy 185 lYsnValLysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuG 205  
 Db 3326 CGGAGCTGGAG-----CTGACGCTGGCGAGCTGG 3355  
 Qy 205 luAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnIleS 225  
 Db 3356 AGCAGGTCGCGAGCTGCTGCCATCTAC---CTGGAGAAGCTCAAGACCATCAGCTGG 3412  
 Qy 225 erLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysT 245  
 Db 3413 TGATCCGCGGCACGACGAGGCGGCTCAGGCGCCACAGGAGCAGCTCAAGG 3472  
 Qy 245 hrThrThrIleValHisAlaThrHisThrArg-----AlaLeuAlaCysIleHisG 262  
 Db 3473 AGGCCAGGCGCTGCCGCGCCACCTCCCGAGCTGAGGCCACCAAGGCTCTCTGAAGA 3532  
 Qy 262 luGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeuLysA 282  
 Db 3533 AGCTGGGGCCCGAGCGGCGAGGCACAGCAGCCACGTTCCAGCCCTCGCGGATGAGCTGC 3592  
 Qy 282 rgAspIleLysLeuValGlySerProLeu---ArgLysTyrSerGluSerMetLysAspL 301  
 Db 3592 GGGGGGCACAGAGGTGGGGGAGGAGCTGCAGCAGCGGCACGGGAGCGGAGCTTGGAGG 3652  
 Qy 301 euSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnG 321  
 Db 3653 TGGAGCGCTGGCGGAGCGGGTCCGCCAGGTGCTTGAGCGCTGGCAGGCTGTGCTGGCCC 3712  
 Qy 321 luMetLysIleArgSerIleGlu-----GlnAsnArgTyrTyrGlnG 335  
 Db 3713 AGACCGACTTGGCGCAGCGCGAGCTCGAGCAACTGGGCGCGCAGCTGCTTACTACCGCG 3772  
 Qy 335 lu-----GlnGlnIleGlnIleL 341  
 Db 3773 AGAGTGCAGACCCCTTGGCGCGCTGCTGCAGGACGCCAGCGCGCGGAGGAGAGATCC 3832  
 Qy 341 ysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGln--- 359  
 Db 3833 AGGCATCCCTGCTGGCCGACAGCCAGCTGTGCGGAGCAGCTGCGGAGGAGCAGCGCC 3892  
 Qy 360 -----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe----- 375  
 Db 3893 TGCTGGAGGAGATCGAGCGCCACGCGGAGAGGTGCGAGAGTGCAGAGGTTTGGAAAC 3952  
 Qy 376 -----AsnLeuHisGluValGlyArgLysSerValLysG 387  
 Db 3953 AGTACATCAAGCCCATCAAGGACTATGAACCTCCAGCTGGTGACGTACAGGGCG---CAGC 4009  
 Qy 387 lySerProValLysAsnProHisLysArg----- 396  
 Db 4010 TTGACCGCGTGGCTCCCGGCCAGAGAGCCCAAGTCCAGTCGGGATCAGAGAGTGTC 4069  
 Qy 397 -----ArgSerHisGluLeuSerPheSerAspAlaProSer----- 408  
 Db 4070 TCCAGGAGTACGTGGACCTGCTGACGCTACAGGAGCTGACCCACCTGACAGCCAGT 4129  
 Qy 409 -----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProL 424  
 Db 4130 ACATCAAGTTTCATCAGCAGGAGCTCGCGCGCATCGAGAGGAGGAGGCTGGCTGAGC 4189  
 Qy 424 ysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerA 444  
 Db 4190 AGCAGCGGCGAGGAGCGCGAGCGCTGGCCGAGGTGGAGCGCGCTGGGAGAGCAGC 4249  
 Qy 444 laArgSerSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaG 464  
 Db 4250 GGCAGCTGGCGCGAGCGCAGCGCCAGGCAAGGCAAGCGCAGCGCGG---GAGCGAAGG 4306  
 Qy 464 laThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrIleG 484

Db 4307 AGCTGCACGACGCGATGACGAGAGAGTGTGTGCGCGGAGGAGGCGCGGTGTGACGCGC 4366  
Qy 484 InProSerValLysLeuTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProA 503  
Db 4367 AGCAGCAGACGCGCAGCATTACGAGGAGAGCTGCAGCAGCTGCGCAGAGCTCGGAGCGG 4426  
Qy 503 spProPheAlaAlaThrThrThrThrLeuAspProPheLeuAlaSerThrMetTyrLeuAspG 523  
Db 4427 AGATCCAGGCCAAGCGCGCAGCAGAGCGGCGTGAAGCGCAGCGCGCTGCATCAGG 4486  
Qy 523 lu-----GlnAlaValAspArgHisGlnAlaAspPheL 534  
Db 4487 AGGAGATCCGCTGCTGCGCTGCGAGTGTGAGGAGCCACCGAGGCCACGCTGCGCGGCGCTG 4546  
Qy 534 yslLysTrpLeuAsnAlaLeuValSerLeuProAlaAspLeuAspAlaAspLeuAsnAsn 553  
Db 4547 AGGGGAGCTGCAGGCATCGCTGTCACCGCGGAGGAGGCTGAGGCACAAAAGCGACAGG 4606  
Qy 554 --LysIleAspValGlyLysLeuPheAsnGluVal----- 564  
Db 4607 CGCAGGAGGAGCGCGAGCGCTTGCAGGAGCGAGGTGCAGGACGAGCGACCGGTAAAGCGCG 4666  
Qy 565 -----ArgAsnLysGluLeuValValAlaProThrLysGluG 577  
Db 4667 AGCGGAGGTGAGCTGCGCTCGCGCTGAAGCGCGAGACCGAGCGCGCGCGGAGGAGC 4726  
Qy 577 luGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThr-----L 591  
Db 4727 AGCGGCGCTCGAGCCCTGAGGAGAGTGTGCGGTGCAGCGGAGGAGCGGAGCGCGCGCC 4786  
Qy 591 euArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys----- 608  
Db 4787 TCGCGCAGCGCGAGTGAAGCGCGGCGGAGGTACAGGTGGCGCTGCAGACGCGCGGAGC 4846  
Qy 609 -----SerLysValAlaValTyrValAsnLysGlnAlaLeuA 621  
Db 4847 GCAGTGCAGAGCGCGAGCTGCAGAGCAAGCGCGCTCTCTCGCGGAGAGACGCGCACAGC 4906  
Qy 621 rgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluL 641  
Db 4907 TGGAGCGCTCCCTGAGGAGGAACACGTGTGTGTGCGACACTCGCGGAGGAGGCTGAG- 4965  
Qy 641 euLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuValValPheGlyGluL 661  
Db 4966 -----CGCGCGCACACGACGACGCGCGCGAGCGCGAGC 4996  
Qy 661 ysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu-----A 679  
Db 4997 GCGCGCGGAGGAGCAGCGCGGAG-----CTGAGCGCTGCGAGCTCAAGGCCA 5047  
Qy 679 snArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrG 699  
Db 5048 ACCAGGCGCTAGGCTGCGGTGCGGCGGAGGAGTGGCGCACAGACAGACCTGCGCGC 5107  
Qy 699 luGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProp 719  
Db 5108 AGCGCGAGGCTGAG----- 5121  
Qy 719 heLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysL 739  
Db 5122 -----AAGCAGAGGAG----- 5133  
Qy 739 ysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaA 759  
Db 5133 ----- 5133  
Qy 759 snIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHis-----A 777  
Db 5134 -----GAGCGGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5185  
Qy 777 rgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgA 797  
Db 5186 GGCAGCGGAGCTGCTGAAACAGAGCTGGAGACGCGCGCGAGCTGCGGAGGAGCGCGCG 5245

Qy 797 spGlyValArgLeuThrArgValValGluValIleLeuLeuArg-----A 812  
Db 5246 CGCAGCAGCGCGCTGCGCGGAGCAGGAGTTGATCCGCTGCGGCGGAGCAGGAGCAGG 5305  
Qy 812 spAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnV 832  
Db 5306 GGGAGCAGCAGCGCGAGCTGTCTGGAGGAGGAGCTGCGCGGCTGCGCGT----- 5355  
Qy 832 alLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaA 852  
Db 5356 -----GAGCGCGCTGCAGCCAGCAGAAACGCGAGAGCTGGAAGCGAGCTGGCC- 5406  
Qy 852 laGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIle 872  
Db 5406 ----- 5406  
Qy 872 yrlLysPheArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArgA 892  
Db 5407 --AGGTGCGGCGCGAG----- 5421  
Qy 892 rgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGluLeuMetArgArgH 912  
Db 5422 -----ATGAGGTGTCTGTGCGCAGCAGCGCGAGGCTGAGGAGAGTGCCTGCTCC- 5472  
Qy 912 isArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValL 932  
Db 5472 ----- 5472  
Qy 932 yslLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgT 952  
Db 5473 -----ACCAGCAGAGTCCAGCAGAGGCTGGAGGCGCGAGCGCGCGGTTCCGCG 5524  
Qy 952 yrlLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrp 972  
Db 5525 AGTGTGCGGAGGCGCGC----- 5544  
Qy 972 tpArgAlaGlnGlnLeuGlyArgGln---HisArgGlnArgPheValGluLeuArgGluA 991  
Db 5545 --CGCTGTGTCCTTGGCGGAGAGGCCAAGCGCGAGCGG-----CAGCTGGCGGAG 5596  
Qy 991 laAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuAlaA 1011  
Db 5597 AAGACGCGCGCGCAGCGCGCGCGAGCGCGGTGCTTCGCGAGAAGCTGCGCGCCA 5656  
Qy 1011 laAlaGlnThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAspTyrIleG 1031  
Db 5657 TCGCGCAGGCGCACCGCGCTC-----AAGACGAGCGCGAGATCGCGCTCAAGG 5704  
Qy 1031 lnMetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnA 1051  
Db 5705 AAGAGGAGCGCGAGACGAGCGCTGCGCGCGCTGCGGAGGAGCGCGCTTCCAGCGCG 5764  
Qy 1051 rgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlnGlnA 1071  
Db 5765 GCGCGCTG-----GAGGAGCAGCGCGCGCACACAAAGCTGACATCGAGGAGC 5812  
Qy 1071 rgLysGluPheGlnGlnLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaA 1091  
Db 5813 GC-----CTGGCCAGCTGCGCAGGCGATCGCAGCGAGCTGGAGCGCGAGAGGCG 5866  
Qy 1091 ysLeu-----SerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgS 1106  
Db 5867 TGTGTGAGGACACCGCTGAGCGCGCGCGAGGTGGAGGAGAGATCTTGGCGCTGAAG 5926  
Qy 1106 eSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgA 1126  
Db 5927 CGAGTTCGAGAGCGCGCGCT---GGCAAGCGAGCTGGAGCTGGAGCTGGAGCGC- 5982  
Qy 1126 snHisTyrTyrSerLeuArgLysAsnVal---IleCysLeuGlnGlnArgLeuArgAlaI 1145  
Db 5983 -----ATCCGCGACGAGCGGAGGAGCAGCTGCGCAGCAGGAGGAGCGCG 6028

1145 leMetIysMetArgGluGlnArg-----GluSerTyrLeuArgLeuArgAsnA 1161  
Db : : : : :  
6029 AGCTGAGGCGCGAGCGACCGCGAGCTGCGCGGAGGAGGAGCGCGCGCGCTGAGG 6088  
QY : : : : :  
1161 laSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181  
Db : : : : :  
6089 CTGAGGAGCGGTGTCAGAGAGCGCTGCGCGCGAGGAGGAGCGCGCGCGAGGAGG 6148  
Db : : : : :  
1181 laTyrIleu-----ArgThrArgLysCysIleIleLeuValGlnArgArgTyrArgA 1198  
QY : : : : :  
6149 CGCGCTGAGGAGAGTGCAGCGCTGAAGCGCTGAAGCGCGAGGAGGCGCGCGCTCGCGG 6208  
Db : : : : :  
1198 laThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysA 1218  
QY : : : : :  
6209 AGCAGCGGAGGAGGAGTGCAGCGCGAGCTGCGCGCGAGGAGGCGCGCGCGAGGAG 6268  
Db : : : : :  
1218 rgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeu 1238  
QY : : : : :  
6269 GGTGTCAGGCGGAGAGAGGAGCACCGCTTCGCGGTGCGAGCAGAGAGGAGGAGGCTAC 6328  
Db : : : : :  
1238 lnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgL 1258  
QY : : : : :  
6329 AG-----CAGACGCTGAGCAGGAGCAGCGCGCTGCGCGAGGAGGAGGAGGAGG 6373  
Db : : : : :  
1258 ysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheH 1278  
QY : : : : :  
6374 GCGAGGCGGAG-----GCGCGCGCGCGCGCGGCGG 6400  
Db : : : : :  
1278 isAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValS 1298  
QY : : : : :  
6401 CTGAGGAGGCGGAGGAGCGCGGTGCGCGAGCGGTGAGGCGCGCGAGGCG----- 6453  
Db : : : : :  
1298 erCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGlnArgAsnSerPheL 1318  
QY : : : : :  
6454 -----CGCGCGAGGTGGAGAGGCGG 6475  
Db : : : : :  
1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetI 1338  
QY : : : : :  
6476 AGCGGTGAAGCAGTGGCA-----G 6496  
Db : : : : :  
1338 leLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGln----- 1355  
QY : : : : :  
6497 AGGAGGAGGAGGAGGCGCGGTGCGCGAGCGGTGCGAGAGCTGCGCAGAGGAGG 6556  
Db : : : : :  
1356 --ThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGln- 1374  
QY : : : : :  
6557 CCGAGCAAGAGGCGCGCGCGCGACAGCGGAGGAGCGCGCGCTGCGCGAGAGAGGAGG 6616  
Db : : : : :  
1375 --ArgGluAlaIleLysValGlnArgArgTyr---ArgGlyAsnLeuGluMetArgL 1393  
QY : : : : :  
6617 CAGCTCAGCGGAGATGGAGAGCATAGAAATTCGCGAGGAGGAGCGCTGCGCGAGAGG 6676  
Db : : : : :  
1393 ysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrTrpA 1413  
QY : : : : :  
6677 CCGAGGTGAG-----CAGAGCTGACACACTGCGGTGAG----- 6714  
Db : : : : :  
1413 rgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerL 1433  
QY : : : : :  
6715 -----CTGAGGAGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6751  
Db : : : : :  
1433 euSerIleGlnArgLysTyrArgAlaThrValGlnAlaArgArgGlnArg----- 1449  
QY : : : : :  
6752 AGGAGTGCAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6811  
Db : : : : :  
1450 -----GluLePheLeuSerThrIleArgLysValArgLeu-----MetGlnAlaPheI 1466  
QY : : : : :  
6812 AGGAGGAGCTTCTTCGCGTGGCGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6871  
Db : : : : :  
1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnArgArgGluP 1479  
QY : : : : :  
6872 TCAGGAGTGAAGACCGGAGCTCATCTGCGTGAGCAAGGAGCAATACGAGCGGCTTCCTGC 6931  
Db : : : : :  
1479 heGluMetLysArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaM 1499

6932 AGGAGGAGGCTGAGAGATGAAGCAGGTGCGGAGGAGGCGCGCGCTGAGTGTGCGG 6991  
Db : : : : :  
1499 etLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgL 1519  
QY : : : : :  
6992 CCCAAGAGGCTTGGCGAGTGGCGAGCTGGCGAGGAGGAGC-----CTGGCAGGAGG 7045  
Db : : : : :  
1519 ysPheArgAlaAsnArgSerMetLys-----GlnAlaArgGlnGluPheValGlnL 1536  
QY : : : : :  
7046 GGGCGTTCGAGAGAGATGCTCAAGGAGAGAGATGTCAGGCGGTGTCAGAGGAGGAGGAG 7105  
Db : : : : :  
1536 euArgThrIleAlaValHisLeuGlnGln-----LysPheArgGlyLysA 1551  
QY : : : : :  
7106 TCAGGCTGAGCGGAGAACTGCTGCGAGCAGCAGAGAGAGCTTGGCAGGAGGAGGAGGAG 7165  
Db : : : : :  
1551 rgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheG 1571  
QY : : : : :  
7166 GGTGTCAGGAGGAGCAGGAGCAGATGCGCAGCAGCTGGCGGAGGAGGAGCAGGAGGCTTC 7225  
Db : : : : :  
1571 lnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGlu- 1590  
QY : : : : :  
7226 AG-----CGGACGCTGGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 7276  
Db : : : : :  
1591 -----MetMetAspLeuIleArg----- 1596  
QY : : : : :  
7277 CTGAGGCGCTCAAGCTGCTGTCGCGCAGATGAGCCGAGGCGCGCGCTGAGGAGG 7336  
Db : : : : :  
1597 -----GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrL 1612  
QY : : : : :  
7337 ACGCCAGCGCTTCCGAGAGCGGCGGAGGAGATCGGTGAGAGCTGCACCGCAGGAGG 7396  
Db : : : : :  
1612 eu----- 1619  
QY : : : : :  
7397 TCGCCACCAGGAGAGGTGACCTGCTGTCGACACTGGAGATCCAGCGAGCAGAGAGTG 7456  
Db : : : : :  
1619 isGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaV 1639  
QY : : : : :  
7457 ACCATGATGCGCGAGCGCTGCGGAGGCGCATCGCTGAGCTGGAGCGTGAAGAGAG- 7512  
Db : : : : :  
1639 alAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleA 1659  
QY : : : : :  
7513 -----AAGCTCCACAGGAGGCGCAAACTGTGTCAG-----CTCA 7546  
Db : : : : :  
1659 laSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrVal-ProHis 1678  
QY : : : : :  
7547 AGTCTGAGGATGTCAGAGCGTGCAGCAGGAGCAGCTGCTGCAGGAGAGCAGGAGGAGG- 7603  
Db : : : : :  
1679 LeuLeuMetTrpCysSerGluPheMetSer-----ThrPheCysTyrGlyIleMet 1695  
QY : : : : :  
7604 -----TGACGAAAGCTTCTCTCTGAAAGGAGCAGCGCTGCTACAGCGGAGGAG 7651  
Db : : : : :  
1696 AlaGlnAlaIleArg 1700  
QY : : : : :  
7652 GCTTCATCGAGCAGG 7666  
Db : : : : :

## RESULT 15

US-09-917-800A-1505  
; Sequence 1505, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880

```
/ PRIOR FILING DATE: 2000-11-02
/ PRIOR APPLICATION NUMBER: US 60/290, 029
/ PRIOR FILING DATE: 2001-05-11
/ PRIOR APPLICATION NUMBER: US 60/290, 645
/ PRIOR FILING DATE: 2001-05-15
/ PRIOR APPLICATION NUMBER: US 60/292, 336
/ PRIOR FILING DATE: 2001-05-22
/ PRIOR APPLICATION NUMBER: US 60/295, 798
/ PRIOR FILING DATE: 2001-06-06
/ PRIOR APPLICATION NUMBER: US 60/297, 457
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/298, 884
/ PRIOR FILING DATE: 2001-06-19
/ PRIOR APPLICATION NUMBER: US 60/303, 459
/ PRIOR FILING DATE: 2001-07-09
/ NUMBER OF SEQ ID NOS: 1740
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1505
/ LENGTH: 15231
/ TYPE: DNA
/ ORGANISM: Rattus norvegicus
/ FEATURE:
/ OTHER INFORMATION: Genbank Accession No. US20020119462A1 X59601
/ US-09-917-800A-1505

Alignment Scores:
Pred. No.: 2,3e-19 Length: 15231
Score: 317,50 Matches: 367
Percent Similarity: 35.11% Conservative: 277
Best Local Similarity: 20.01% Mismatches: 670
Query Match: 3.34% Indels: 522
DB: 81 Gaps: 81

US-09-914-698-1 (1-1861) x US-09-917-800A-1505 (1-15231)
QY 52 LysThrLeuGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValSer 71
Db 3565 AAAGTTCTGCTTGGCAGAGCGTGCACCTGTCACCAACTCTGCGCTGGAGTTGAA 3624
QY 72 AlaAlaValGlnGlnLysLys-ArgMetSerAlaAlaAlaAlaProSerLysGlnTh 91
Db 3625 TTGACCTGGGCAAGCTGGGAACAGG---TCAGAAGCCTGTCTGCCA-----TCTAC 3672
QY 91 rTPArqValThAlaProSerArgProAlaAlaTTPAlaAlaHisProProGlnAlaPr 111
Db 3673 TTGGAGA-----AACTCAAGACCATCAGCTTGTATTTGCAGTACCCAGGGGGCT 3723
QY 111 oLeuValGluLysAsnValTyrLysThrProGlnGluGluProValTyrLysSerProG1 131
Db 3724 -----GAGGAGGTGCTTAAACACACAGGAGGACACCTGAAGGAGGCCCGCCGCT 3773
QY 131 nProArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspVal-- 150
Db 3774 GCCTGCCACACTCCAAG-----CTCGAAGTCAAC 3803
QY 151 -----IleAspAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyG1 167
Db 3804 CAAGGCTTCACTAAGAGAGTGGCG-----GCCAGCGGAGGCACA 3845
QY 167 nAlaThrIlePheProAspAsnLeuAlaAlaTTPProThrProThrLysGlyAsnVa 187
Db 3846 GCAGCCTGTATTC----- 3858
QY 187 LysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuGluAspG1 207
Db 3859 -----AACACCTACGA-----GATGAGCTGAGGGGGC 3887
QY 207 nProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnLysSerLeuAs 227
Db 3888 ACAGGAGTTGGTAACGGCTACAGCAGCGCGCATGGTGAGCGG----- 3930
QY 227 pThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThr 247
Db 3930 ----- 247
```

```
Db 3931 -GACGTGGAACTAGAGCGCTGGCGAGAACGTGTCTCACTGCTTGTGGAGCCCTGGCAGCC 3989
QY 247 rIleValHisAlaThrHisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSe 267
Db 3990 TGTCTAGCCAGACTGATGTGCG----- 4014
QY 267 rProProArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuVa 287
Db 4015 -----CAGCGGAGCTTGAACAGCT 4034
QY 287 lGlySerProLeuArgLysTyrSerGluSerMetLysAspLeuSerLeuLeuSerProG1 307
Db 4035 GGGCGGCCAACTTGGCTACTACCTGAAAGTGGCGGATCCGCTGAGCTCCTGGCTGCAGGA 4094
QY 307 nThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerIl 327
Db 4095 TGCCCAAG----- 4101
QY 327 eGluGlnAsnArgTyrTyrGlnGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSe 347
Db 4102 -----AGCCGCGCAAGAACAGATCCAGCTGTGCCAATAGCCAA 4139
QY 347 rSerSerSerSerGluAlaSerLeuAlaGlyGlnGln-----GluPheLeuPheAs 364
Db 4140 CAGTCAGGCTGCACGAGAACAGCTGCCAGGAGAGGCCCTGCTGGAGGAGATTGAGCG 4199
QY 364 nHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGlyArgLysSe 384
Db 4200 CCATGTTGAGAAGGTTGAGGAGTGCCAGAAGTTTGTCTAAGCAGTACATC-----AATGC 4253
QY 384 rValLys-----GlySe 388
Db 4254 AATCAGGACTATGAGCTCCAGCTGATCACTCAAGGCTCAGCTTGAAOCTGTGGGCTC 4313
QY 388 rProValLysAsnProHisLysArgArgSerHisGluLeuSerPheSerAspAlaProSe 408
Db 4314 CCGCGCCAGAGCCCAAGTTTCAGTGTGATCGGAGAGGCTCATCCAGAGTACGTGGA 4373
QY 408 rAsnGluSerLeuTyrArgAsnGluThrValAlaIleSerPro----- 422
Db 4374 TCTGCTGTACAGCTACAGTGTGAGTGCACCACTCACTCAGAGTCAGTACATCAAGTTTCATCAG 4433
QY 423 -----ProLysLysGlnArgValGluAs 430
Db 4434 TGAGACACTGCGCCGCGATGGAGAGGAGCGGCTGCTGAGCAACAGCGGCGAGGGA 4493
QY 430 p---ThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerAl 449
Db 4494 GCGGAGCGGCTGGCCGAGGTGGAGCGCGCTGGAGAGCAGCGGCGAGCTGCTGAGGC 4553
QY 449 aHisAlaTTPProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMe 469
Db 4554 CCATGCCAGCGCCAGAGGCACAGGCC---GAGCTGGAGGCGACGAGAACTGTCAGCGGGCAT 4610
QY 469 tLysLysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLysLe 489
Db 4611 GCAGGAGAGGTGACCGCGCGGAGGAGCGCGGTGGACGACACAGAAAGCGCAG 4670
QY 489 uTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProAspProPheAlaAlaTh 508
Db 4671 CATCCAAGAGGAGCTGACGATCTGGCGCAAGCTCAGAGGCGAGAGATCCAGCCCAAGGC 4730
QY 508 rThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGlu----- 523
Db 4731 CCAGCAGGTGGAGCTGTCAGAGCGCGCATCGCATTTAGAAAGAGATCCGCGTAGT 4790
QY 524 -----GlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAl 539
Db 4791 CCGTCTGAGCTAGAGCAACACTGAGCTGAGCTGGAGGGGCGGAGGATGAGTCAGGC 4850
QY 539 aLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn---LysIleAspValG1 558
Db 4851 TCTGCTGTACGGCTGAGGAGGCGAGACACAGAAAGCGGCGAGCTCAGGAGGAGCCGA 4910
```



```

Db      6804  |||:||||: ||| ||| |||:||||: :
      GGGCCAGCGCAGCGCGGTGCAGAGAACTCGCAAGAAAGCGGAGGAGCGCGCGC 6863
Qy      1190  leAsnValGln-----ArgArgTrpA 1197
Db      6864  TCGGGCCAGCGCGAGCAGCGGTGGTGAACACAGAGCAGGACGCCACCGGAGATGGA 6923
Qy      1197  rgAlaThr-----LeuGlnMetArgArgGluArg---LysAsnTyrLeuHisLeuGlnT 1214
Db      6924  GAAGCACAAAGATTTCAGAGCAGACAGCTA-CGGCAGAAAGCTCAGGTAGACGAGCAGC 6982
Qy      1214  hrThrThrLysArgIleGlnleLysPheArgAlaLysArgGluMetLysLysGlnArgA 1234
Db      6983  TGACCACGCTGAGGCTGCAGCTCGAG-----GAGCCGACCACCCAGGAAGA 7027
Qy      1234  la-----GluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgA 1251
Db      7028  GCATCTCGATGAGGAGCTGCAGCGCTAAAGGCTGAGGTAAACAGAGAGCGCCCGGCAGC 7087
Qy      1251  rgAlaLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArg-----G 1268
Db      7088  GTAGCCAGGTAGAG-----GAGGAGCTTCTCTGTCTCGCGTGCAGATGG 7132
Qy      1268  luValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgA 1288
Db      7133  AGGAGCTGGCAACTCAAGCTCGATTGAGTGAACCCGGCGACTCATCTTCGTG 7192
Qy      1288  laLysTyrArgGlyThrGlnAlaValSerCysLeuGlnMetHisTrpArgAsnHisL 1308
Db      7193  ACAAG--GACAACACACAG-----CGCTTCC 7216
Qy      1308  euLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuG 1328
Db      7217  TGGAGGAGGAGCGCCGAGAAGATGAAACAGGTGGCAGAGAGAGTGCACGGTTGAGCGTAG 7276
Qy      1328  lnArgArgTyrArgAlaArgLeuAsnMetIle----- 1338
Db      7277  CTGCCAGCAGCAGCAAGGCTGGCGCAGCTAGCGGAGGAGCCTGGCCCGCAGCAGCGGG 7336
Qy      1339  -----LysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleG 1355
Db      7337  CCTGGCGGAGAGATGCTGAAGAGAGAGATGCAAGCGCGGTGCAGGAAGCC----- 7386
Qy      1355  lnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnA 1375
Db      7387  --ACAGGCTCAAGCT-----GAGGCTGAGCTGCTGCAGCAGCAGA 7426
Qy      1375  rgGluAlaIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnI 1395
Db      7427  AGGAGCTGGCAGCAGGAGCGCGCGCTGCAGCGGACAGAG----- 7473
Qy      1395  leGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerI 1415
Db      7474  -----CAATGGCTCAGCAGTGGTAGGAGACACAGGGTTTCAGCGGACCC 7522
Qy      1415  le-----ArgAspMetArgLeuCysLysAlaLysLysArg---ArgIleA 1429
Db      7523  TGGAGGCTGAGCGCGCAGCGCAGCTAGAAATCAGCGCAGAGGCTGAACGCTCAAGTTGC 7582
Qy      1429  rgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGlnA 1449
Db      7583  GCATGGCTGAGATGAGC---CGGGCTCAGCGCGCTGCAGAGGAGGTGCCCGCGCTTCC 7639
Qy      1449  rgGlu-----IlePheLeuSerThrIleArgL 1458
Db      7640  GGAAGCAGGCTCAAGAGATCGCGAAAGCTGCACCGCAGCTGAACCTGCTACACAGGAGA 7699
Qy      1458  ysValArgLeuMetGlnAlaPheIleArgAlaThrLeu---LeuMetArgGlnGlnArg- 1476
Db      7700  AGGTGACATTTGTCAG-----ACTCTCAGATCCAGCGCAGCAGAGATG 7744
Qy      1477  --ArgGluPheGluMetLysArgAlaAlaValIleGlnArg----- 1491

```

```

Db      7745  ACCAAGATCCGAGCGTCTGACGGAGGCCATTGCTGAOCTGGAGCGTGAGAGAGAGAC 7804
Qy      1492  --ArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrTrpArgGlns 1511
Db      7805  TCAAGCAGGAGCGAAGTTACTGCGAGCTCAAGTCT--GAGGAGATGCAGACTGTGCAGC 7861
Qy      1511  erSerValIleLeu-----ValGlnArgLysPheArgAlaAsnArgSerM 1526
Db      7862  AGGAGCAGATACTGCGAGGAGACACAGGCCCTTCAGAAAGAGCTTTCTCTGAAAGAGCA 7921
Qy      1526  euLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnL 1546
Db      7922  GCTTGTGTCACCGAAGCGCTTCATCGAGCAGAG-----AAGGCCAAGCTGGAGCAGC 7975
Qy      1546  ysPhe-----ArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheG 1561
Db      7976  TTTTCCAGGACGAGGTGGCAAAAGCAAAACAGCTGCAGGAGGAGCAG----- 8022
Qy      1561  lnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgL 1581
Db      8023  -----C 8023
Qy      1581  ysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIle----- 1595
Db      8024  AGCGGCAGCAGCAGCAGATGGAGCAGGAGAAACAGCAGGAGCTGGTGGCCAGCATGGAGGAG 8083
Qy      1596  -----ArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyr----- 1611
Db      8084  CCGGAGCGCGCAGCGTGGAGCAGAGAGGGTGTGAGGCGCAAGCAAGAGAACTGCAGC 8143
Qy      1612  --LeuIleArgArgArgGlnLysHisGlnGlyLeuLeu-----AspIleA 1626
Db      8144  GTCTGAGCAGCAGCAGCGCAGCAGCAGGAGAACTACTGCGAGGAGAACACAGAGGCTGC 8203
Qy      1626  rgLysArgIleAlaGlnLeuArgGlnGluAlaLysAla 1638
Db      8204  GGGAGCGCTGCAGCGCTTCGAGGAAGAGCACCGAGCT 8241

```

Search completed: September 15, 2004, 13:45:43  
Job time : 2211 secs

*This Page Blank (usp10)*